

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 - 703-308-4498
ian.delaval@uspto.gov

STAFF USE ONLY

Type of Search


Vendors and cost where applicable

Searcher: <u>Jan</u>	NA Sequence (#) _____	STN _____
Searcher Phone #: <u>4498</u>	AA Sequence (#) <input checked="" type="checkbox"/>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>6/6/02</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>6/6/02</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <input checked="" type="checkbox"/>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____


This Page Blank (uspto)

68219

Delaval, Jan

From: 
Sent: Thursday, June 06, 2002 9:58 AM
To: Delaval, Jan
Subject: RE: 09/927,458

Jan,

 Please search polypeptide of SEQ ID NO: 2 against commercial and interference databases.

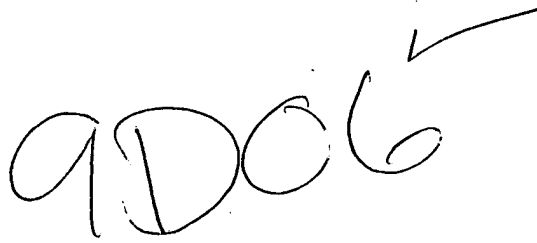
Thanks,

Neon

Art unit 1644 ✓

Mail 9E12 ✓

Tel 308-4844



Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 - 703-308-4498
jan.delaval@uspto.gov

This Page Blank (uspto)

C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T46366
 R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, January 2000
 A:Reference number: Z23031
 A:Accession: T46366
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-547 <AAA>
 A:Cross-references: EMBL:AL137520
 A:Experimental source: adult testis; clone DKFzp434C0118
 C:Genetics:
 A:Note: DKFzp434C0118.1

Query Match 5.3%; Score 145.5; DB 2; Length 547;
 Best Local Similarity 21.7%; Pred. No. 0.0085;
 Matches 98; Conservative 61; Mismatches 123; Indels 169; Gaps 24;
 QY 105 SLATSSNOTSMQDEQPSDSFQQAAGSVMDDMLG-PSQNFEEASIQDNNHMAEGTGF 163
 Db 72 SSAATGSRHRPDTQHPSS--GGRCRGGTSPSSAAGRPASMAEAE---EDCH----- 119
 QY 164 YPSEPMLCSESVGQVPHSLSTLYQSDCSDANDALIVLHLLMLESYGYPQGTAKA-L 222
 Db 120 -----SDTVR-----ADDEENES-----PAETDLQAOL 143
 QY 223 SMPEKWLKSGYKIQVMHPLCEG-SSATLTCVPL-----GNLVNATLKNNE----- 270
 Db 144 QM-----FRAQWFMELAPGVSSSNLENRCPRAAGSLQKTSADTKGQKQAEKVA 194
 QY 271 -----IRSV-----KRLQLPE-----SFICKEKLGENVANIY-----KT 302
 Db 195 RELFLKAVEEQNGALYEAIFKFRAMQLVPDIEFKTYTTRSPDGVGNSYIEDND 254
 QY 303 KLSRL---FKDQLVYPLIAFTRQALNL--PDV-----FGLVVLPLELKLRLFR 348
 Db 255 KMADLLSYFQQQ-----LTFOESVLKLCQPELESSQIHISVLPMEVLMIYIFRWVSSDLD 309
 QY 349 VRSVLSLAVCRDLFTASNDPLLRFLVLRDFRNTVR-VQDITDKELYRKRHIQRKESP 407
 Db 310 LRSLEQLSLVCRGFTCARDPEIWRACLKLVWGRSCIKLVPTYSWRMFLEL-----P 362
 QY 408 KGRFVMLPFSST-----HTIPYFNPPLHPRPPFPSSRLPGIIGGEYDQ 450
 Db 363 RVRFEGVVISKTTYIROGEQSLDGFYRAWHQVEYY---RYIRFPDGHV----- 408
 QY 451 RPTLPYVGDPISSLIPGPGTSPQPPPLRPR 481
 Db 409 -----MMLTTPPEPQSIIVPLRTR 427

RESULT 4
 T32258
 hypothetical protein C24A1.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32258
 R:Connell, M.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid C24A1.
 A:Reference number: Z21141
 A:Accession: T32258
 A:Molecule type: DNA
 A:Residues: 1-842 <CON>
 A:Cross-references: EMBL:AF024491; PIDN:AA070312.1; GSPDB:GN00021; CESP:C24A1.3
 A:Experimental source: strain Bristol N2; clone C24A1
 C:Genetics:
 A:Gene: CESP:C24A1.3
 A:Map position: 3
 A:Introns: 43/2; 189/3; 234/1; 351/1; 511/3; 578/3; 594/2; 636/2; 675/3

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
 C:Accession: H86371
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dwar, K.;
 ansen, N.F.; Hughes, B.; Hultzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: H86371
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-350 <STO>
 A:Cross-references: GB:AE005172; NID:94056459; PIDN:AAC98032.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 6.8%; Score 187.5; DB 2; Length 350;
 Best Local Similarity 24.4%; Pred. No. 3.7e-06;
 Matches 90; Conservative 58; Mismatches 118; Indels 103; Gaps 17;
 QY 172 SESVGEQVPHSL-----TLQSDCSDANDALIVLHLLMLESYI 213
 Db 22 NSGIEGVPMDVLEAAAKSKLSEFFLKNVLLERKSGTSDLT-ALALSVAHVAWLESGFV 80
 QY 214 --PQGT-----AKALSMPEKWLKSGYKQVMHPLCEGSSATITCVPLGNLIVNATL- 265
 Db 81 LLDHSGDKFSFKLLSVSLRYTLPELITRKDTNTV---ESVTVRFQNGIRFLVYGTG 137
 QY 266 ----KTNIRSVKRLQLLPESFICKEL---GENVANIYKDLKSLRFLKDLVYPLLA 318
 Db 138 GSKCRVH--MTSLDKSRFLPVLDLVVDLTKFKEGSSSYREVFMWVWVWVDELVIFLL- 194
 QY 319 FTQALNLPDVF-----LVVLPLELKLIRFLRLDVRSLVSACVCRDLFTASNDPLW 372
 Db 195 -----IGLCKAGLESPPCLMLPTELKILLELPVGSIGVACVCTEMRYLASDNDLW 249
 QY 373 RFLYLRDFRNTVRVO--DTWKELY-----RKR--HQIKESPKGREVMLLPSSHTIP 423
 Db 250 EHKLEEGKGLKWLKTVGDVWKRKFASFWRKRRLDLLARRNPPIK----- 296
 QY 424 FYENPLHPRFPSSRLPPGIIGGEYDORPTLPVGDPISSLIPGPGTSPQPPPL----- 478
 Db 297 --SNRPFPTLFPDR-----DRPEPDRFG-----PSDFYRFLRDP 331
 QY 479 RPRFDPVGP 487
 Db 332 RDRFGPRDP 340

RESULT 3
 T46366
 hypothetical protein DKFzp434C0118.1 - human (fragment)

Query Match 4.7%; Score 128.5; DB 2; Length 842;
Best Local Similarity 20.3%; Pred. No. 0.29;
Matches 109; Conservative 79; Mismatches 197; Indels 151; Gaps 25;

QY 14 LEVPEETPTGHLRSHLRLSLCTWGYSSNTRFTITLNYKDPDLGDEETLASYGIVSGDL 73
DB 334 MKYGVDPNVOGDGHTALHSACYVHGLRIVQYLLE-----NGAQSLASRAFEAGAL 386
QY 74 ICLILQDDIPAPNIPSTSDSHSLQNNQPSLATSNQTSNOTSMQDEQPSDSFGQQAASGV 133
DB 387 ----RQAGPCTNRPKSVASAIMALNSDTPSSNASTNSTVSLDDQQTVP-----I 433
QY 134 W-----NDDSLMGPSONFAESIODNAHMAEGTFYFSEPMCLSCSEVQ-----VPHSLE 184
DB 434 WAYERGHDAIVALLKHVAARTVEGD-----VCSEYSSGESSYTPLPSPMG 478
QY 185 TL-----YQADCSANDALIVLIHLLMESY----IPQGTAKALSMPKWKLSGVYKLO 237
DB 479 RLTSLTRDKADLLQRLSALPAPFHLCAETEFQESIGSGFGK-----VYKGT 526
QY 238 YMHPLCEGSSATLRCVPLGNLIVVATLKINNEIRSVKRLQLLPESPTCKEKLGENVANI 297
DB 527 YR-----GKLV-----AVKRRAM--AFCKSE-----TDM 551
QY 298 YKDLQKLSRLFKQOLVPLLAFTQALNLPDVLGVLPLEL-----KLRIFRLLD 348
DB 552 CREVSLSLRLAHNVV----AFVGTSLDDPSQFALITEFVENGSLFRRENGERKNYRMD 607
QY 349 ----VRSVLSLSAVCRDLFTASNDPLLWFL-----YLRDRDNTVRVQD 391
DB 608 PAFRLRLSLDVARGMRYLHESAAPVTHRDNLNSHNLIHADGRSVWADFGESREVCQRED 667
QY 392 WKELYRKHRIORKEPK-----GREVMLPSTHTIPFPNPLHPRFPSPRLPGIIG 446
DB 668 -ENLTQPGNLRWMAPEVFSOGYDKRVDVFSALVIW--EHTAELPFSHLKAPAAAA 724
QY 447 E-YDQ-RPTLPYVGD-----ISSLPQGETPQFPPLRFRFPVGPPLGPN 492
DB 725 EMTYKRGKRTLP--NQTAQFPAHLSLIQAWHPSS---SRPDFVEIVALLEPH 775

RESULT 5

H70804
hypothetical protein Rv3494c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70804
R:Cooper, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70804
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-564 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17731.1; PID:e125462
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv3494c

Query Match 4.3%; Score 118; DB 2; Length 564;
Best Local Similarity 25.5%; Pred. No. 0.93;
Matches 51; Conservative 18; Mismatches 51; Indels 80; Gaps 11;

QY 354 SLSAVCRDLF--TASNDPLLWFLYLRDRDNTVR-VQDITDKELYRKR--HIQKESPK 408
DB 343 SVREIPDMYCKTAQNDP-----STVRGARNYPCQFPKGRAPTVOLCRDRP 389

QY 409 GRFVMLLPSSSTH-----TIPFP-----NPLHPRPFPSSRLPPGIIGGEYDORPTLPV 457
DB 390 G-----YVPVGTNPNRGPPIPGTGVTEVTDGRNLLPNKFP--YIPGCA-----DPDGGVPIV 438
QY 458 GDPISLILPGGEPSPQ-----FPLRLPRF----- 482
DB 439 GPPPGGVAGPGAPHAPQAPPPNDNGPPPTTSWMPGYPPEPPQVPYPATIPPPPP 498
QY 483 ----DPVGLPLGPNLILPG 497
DB 499 PEGTGPFPAPGAPGPOQASG 518

RESULT 6

A41724
limb deformity (ld) protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 04-Mar-1993 #sequence_revision 15-Aug-1997 #text_change 10-Sep-1997
C:Accession: S24286; S38780; A41724
R:Trump, A.; Blundell, P.A.; de la Pompa, J.L.; Zeller, R.
Genes Dev. 6, 14-28, 1992
A:Title: The chicken limb deformity gene encodes nuclear proteins expressed in specific limb bud tissues
A:Reference number: A41724; MUID:92112031
A:Accession: S24286
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1213 <TRU>
A:Cross-references: EMBL:X62681
A:Experimental source: embryo
R:Zeller, R.
submitted to the EMBL Data Library, August 1991
A:Reference number: S38780
A:Accession: S38780
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-885, 'A', 887-1213 <ZEL>
A:Cross-references: EMBL:X62681; NID:g63567; PID:g63568
C:Comment: Mutations in this gene affect morphogenesis of both limbs and kidneys.
C:Genetics:
A:Gene: ld
C:Keywords: nucleus

Query Match 4.3%; Score 117.5; DB 2; Length 1213;
Best Local Similarity 19.4%; Pred. No. 3.1;
Matches 105; Conservative 55; Mismatches 204; Indels 177; Gaps 19;

QY 57 TGDEETIASYGVSGDLICLILQDDIPAPNIPSTSDSEHSLQNNQPSLATSN----- 111
DB 297 TEGGETITEIKPRENDLALLKRPVKNSITSLGTTKKRSSEKSPKPTFLEQLSHLNI 356
QY 112 QTSWDEQPSDSFGQAQSGVWDDSLMGPSONFAESIODNAHMAEGTFYFSEPMCL 171
DB 357 DVSNDERTQDSGAGFE-----TEDSGEPENKASQ-----TEFLFP 395
QY 172 SESYGVQPHSLTYQSADCSANDALIVLIHLL-----MLESYIIPQTEA-KALSMP 226
DB 396 SEEEKS-----SPAESALDVKALFTTRPKKETTADPSELEAIKRRNE 440
QY 227 KWKLGSYVLYKYMPLCEGSSATLTCVPL-----GNLIVVATLKINNEIRSVKRLQL 279
DB 441 KESLKAVERFSKSKPGDPSDKSPSEQDDKTPGRLOTVPWPPPKRANHE----- 491
QY 280 LPESFICKELGENVANIYKDLQKLSRLFKDOLYVPLLAFTQALNLPDVLGVLPLEL 339
DB 492 ----EVKVLKYTEAEYQAAILHLKREHKE-----IETLSQF 526
QY 340 KLRIFRLLDVRSLSL-----SAVCRDLFTASND-----PLLWRF 376
DB 527 ELRVFHNGEHAHVSTAQLEETIAHLKNELDKNLNRNEARDIGVSTEDDNLPKYRNC 586
QY 377 LRDRDNTVRVQDITDKELYRKHRIORKEPKGRFVMLLPSSSTHTI----- 422

Thu Jun 6 11:44:35 2002

```
Db 587 IOTDRETFIKPSEENRAVNQIIVPKK-----LNISLTHSIQGENKSDYDVS 638
Qy 423 -----PYPNPLHPRFPSSRLPPGIIGGEYDQRTLPYVGDPISSLIP-- 466
Db 639 SESVLCQKQMLPSPPPPPPPPPPP-----PP-----PPPSDSSLPLCLVPPP 684
Qy 467 -----GPGETPSQF---PPLRPRF-----DPVGLPGNPNLPLRGNGNDPRFPSPGR 513
Db 685 PPLPTGTSVTHFARGPPLPOLSEGRDFOAPAPPAPPLPLGLGPPVP--PPLPGSGL 742
Qy 514 P 514
Db 743 P 743

RESULT 7
S52796
prpL2 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997
C:Accession: S52796
R:Ruhmann, A.; Kreideweiss, S.; Nordheim, A.
Submitted to the EMBL Data Library, March 1995
A:Reference number: S52796
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-403 <RUH>
A:Cross-references: EMBL:X86019; NID:g762950; PID:g762951

Query Match 4.2%; Score 117; DB 2; Length 403;
Best Local Similarity 32.9%; Pred. No. 0.67; Indels 44; Gaps 9;
Matches 46; Conservative 9; Mismatches 41;

Qy 415 LPSSTHTIPFPNPLHPRFPSSR-----LPPGIIGGEYDQRTLPYVGDPISSL---IP 466
Db 203 VPSYPR--PSAPHRPLRLPPPPSPRPGPPPLPPSSSGN--DETRLPQRNLSLSSTPTPLP 258
Qy 467 GPGET-----PSQFPPLRPRDP--VGDLPGNPI-----LPG----- 497
Db 259 SPGRSGPLPPVPVPSRPPPPVDRDPPGRSGPLPPPPVSRNGSTRALPATPQLPSRSGVD 318
Qy 498 --RGGPNDRF--PFRPSGRP 514
Db 319 SPRSGPRPPPLPPDRPSAGAP 338

RESULT 8
T12526
hypothetical protein DKFZp434M183.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C:Accession: T12526
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
Submitted to the Protein Sequence Database, June 1999
A:Reference number: T12524
A:Accession: T12526
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-915 <WAM>
A:Cross-references: EMBL:AL080141
A:Experimental source: adult testis; clone DKFZp434M183
C:Genetics:
A:Note: DKFZp434M183.1

Query Match 4.2%; Score 116.5; DB 2; Length 915;
Best Local Similarity 20.7%; Pred. No. 2.4; Indels 209; Gaps 36;
Matches 129; Conservative 76; Mismatches 208;

Qy 34 LLCTWGYSSNTRTITLNY-KDPLTGDEETLASYGIVSGDLICLILODDIPAPNIPSTSD 92
```

```
Db 205 LKVTLEQDSRMKFLKLGYSKDEL---OKKVATW-----LKSDV---GLGESQ 247
Qy 93 SEHSSLONNEQPSLATSSNOTSMQDEQPSDSFGQAQSGV--W-----NDDSM----- 139
Db 248 PKGNDLNSDRQQAFCQASKHTTKEASAFFDELVPQNMTPEIPITKIDGLLSQAL 307
Qy 140 ----LGPSON--FEAESIODNAHMAEGTGF----YPSEPMLCSESVGQVPHSLLETLYQS 189
Db 308 LLGELGPAVELCLKEERFADAIILAQAGGTDLKQOERYLAKKKT--KISSLLACVVGK 365
Qy 190 -----ADCSAN--DALIVLIHLLMESGYIPQGTAKALSMPKWKLSGVYKLYMHP 241
Db 366 NWKDVVCTCSLKNWREALALLTY-----SGTE---KPELCLDMLGTRMEQ---- 408
Qy 242 LCEGSSATLTCVPLGNLIVVNATLKINNEIRSVKRLQLLPESFICK---EKLGENVANIY 298
Db 409 --EGSRA-----LTSEARL-----CYVCSGSVERLVECWAKCH 439
Qy 299 KDQLKSLRLEKFDQIVYPLLAFTQALNLPDVFGLVVLPLELKIRIFRLLDVRSVL----- 353
Db 440 --QALSPMALQDLMEKVMVNLRSLEQLRGPHGVSGP-----ATTYRVTOYANLLAAQGS 492
Qy 354 ---SLSAVCRD-----LFTASNDPLLR-----FLYLDRDFRDNTRVQDQDWKE 394
Db 493 LATAMSLPRDCAQPPVOQLRDLFHAQGSVILGQQQPPFPFRIYVGVTLHSKETSSYR 552
Qy 395 L-YRKRHIORKESPKGRFVMLLPSTHTITFYFNPPLHPRFPSSR----- 438
Db 553 LGSQPSHQVPTPSRPR--VFTPOSSPAMPPLAPS--HPSPYQGPRTQNIISDYRAPGQAI 608
Qy 439 ----LPPGI-----IGGEYDQRP-----TLPYVGDPISSLIPG---PG----- 470
Db 609 QPLPLSPGVRPASPQQLLGGQVQVQVNPVGFPGTWPPLPGSPLMACPGIMRPGSTSL 668
Qy 471 TPOSP--PLRP-----RDPVGLPGPNPILPG-----RGG--D 503
Db 669 TPRLEPLPLPLRGPRMVSHTPAPPASFPVPLPGDGPAPCSSVLPTTGILTPHPGPQD 728
Qy 504 RPPFRPS--RGRTDQRL--SPM 522
Db 729 SWKEAPAPRGNLQRNKLPETEM 750

RESULT 9
T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya
Oncogene 15, 549-560, 1997
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A:Reference number: 214954; MUID:97388474
A:Accession: T03455
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

Query Match 4.2%; Score 116; DB 2; Length 4957;
Best Local Similarity 20.0%; Pred. No. 33; Indels 164; Gaps 25;
Matches 102; Conservative 64; Mismatches 181;

Qy 94 EHSSLONNEQPSLATSSNOTSMQDEQPSDSFGQAQSGV--GVWNDD-SMLGSPSQNFEAS 150
Db 3356 QQQQLQQQQQQQLQQQQQQQLQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 3409
```


Search completed: June 6, 2002, 11:30:25
Job time: 159 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM-protein - protein search, using sw model

Run on: June 6, 2002, 11:24:26 ; Search time 34.41 Seconds
(without alignments)
1684.990 Million cell updates/sec

Title: US-09-927-458-2
Perfect score: 2754
Sequence: 1 MLRLVLLKRWLPVETPE.....DRFPFRPSGRPTDGLRSLFM 522

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneq/geneq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2744	99.6	522	22	AA032108
2	2510	91.1	591	19	AA068521
3	2489	90.4	482	21	AA083047
4	2448.5	88.9	607	22	AA032109
5	2270	82.4	443	22	AA035160
6	2086.5	75.8	549	22	AB018510
7	714	25.9	175	20	AA041397
8	502	18.2	94	22	AA032107
9	299	10.9	113	21	AA087356
10	265.5	9.6	225	22	AB06838
11	264.5	9.6	53	22	AB018509

12	264.5	9.6	53	22	AA032108	Novel human secret
13	240.5	8.7	475	21	AA029404	Arabidopsis thalia
14	194	7.0	39	21	AA083054	F-box motif of FBP
15	189	6.9	38	20	AA020272	A F-box protein se
16	189	6.9	38	22	AA080844	Human F-box protei
17	164	6.0	38	20	AA020273	A F-box protein se
18	164	6.0	38	22	AA080845	Mouse F-box protei
19	152.5	5.5	336	21	AA029406	Arabidopsis thalia
20	152.5	5.5	342	21	AA029405	Arabidopsis thalia
21	147	5.3	270	22	AB064135	Drosophila melanog
22	145.5	5.3	430	22	AA041539	Human polypeptide
23	144	5.2	321	21	AA092341	Human cancer assoc
24	144	5.2	405	22	AA039753	Human polypeptide
25	144	5.2	437	22	AA093745	Human protein sequ
26	139	5.0	327	21	AA083092	F-box protein FBP-
27	139	5.0	327	22	AA035162	Human Skpl-associa
28	139	5.0	339	22	AA092278	Human digestive sy
29	127.5	4.6	712	22	AB018031	Novel human diago
30	126.5	4.6	965	22	AB023404	Novel human diago
31	126	4.6	355	22	AB063369	Amino acid sequenc
32	126	4.6	355	22	AA036629	Human FLEXHT-51 pr
33	123	4.5	271	20	AA031376	Human proteasome-1
34	121.5	4.4	195	22	AA030711	Novel human secret
35	120.5	4.4	2091	21	AA012000	Rat p3103 protein.
36	118	4.3	442	22	AB071447	Drosophila melanog
37	117	4.2	504	20	AA024091	Human Wiskott-Aldr
38	116.5	4.2	916	22	AB009870	Novel human diago
39	116.5	4.2	969	22	AA000937	Human bone marrow
40	116	4.2	966	22	AA000988	Human bone marrow
41	114	4.1	245	22	AA036620	Human FLEXHT-42 pr
42	114	4.1	265	22	AB019238	Novel human diago
43	111.5	4.0	940	22	AA028194	Novel human secret
44	111.5	4.0	968	22	AA028382	Novel human secret
45	110	4.0	314	22	AA01594	Human gene 12 enco

ALIGNMENTS

RESULT 1

AA035161

ID AA035161 standard; Protein; 522 AA.

AC AA035161;

XX 09-APR-2001 (first entry)

DT Human Skpl-associated F-box protein-1 SAF-1beta SEQ ID NO: 10.

XX Human; protein degradation; siah-mediated degradation protein; SMPD;

XX SCF-complex protein; SCP; siah-1alpha; siah-1 interacting protein; SIP;

XX Skpl-associated F-box protein; SAF-1; SAF-2; SAD; cancer; cell division;

XX Skpl-associated destruction-box protein; inflammatory disease.

XX Homo sapiens.

XX WO200077207-A2.

XX 21-DEC-2000.

XX 09-JUN-2000; 2000WO-US15873.

XX 11-JUN-1999; 99US-0330517.

XX (BURN-) BURNHAM INST.

XX Reed JC, Matsuzawa S;

XX WPI; 2001-071273/08.

XX N-PSDB; AAC67285.

XX Siiah-Mediated Degradation Protein, useful for drug screening, for

XX Therapeutic applications and for functional genomics -

XX PS Claim 15; Page 107-108; 121pp; English.

XX CC The present invention provides the protein and coding sequences of

XX CC several shah-mediated degradation proteins and SCF-complex proteins.

XX CC These are designated shah-1alpha, shah-1 interacting protein (SIP), which

XX CC encodes two proteins due to alternative splicing (SIP-L and SIP-S),

XX CC Skp1-associated F-box protein-1alpha and beta and -2 (SAF-1alpha,

XX CC SAF-1beta and SAF-2) and Skp1-associated destruction-box protein (SAD).

XX CC The proteins and their coding sequences are useful in the diagnosis and

XX CC treatment of cancers, disorders where too little cell division occurs

XX CC such as bone marrow aplasia, immunodeficiencies and inflammatory

XX CC diseases including sepsis, fibrosis, arthritis and graft versus host

XX CC disease.

XX CC Sequence 522 AA;

XX CC

XX CC Query Match 99.6%; Score 2744; DB 22; Length 522;

XX CC Best Local Similarity 99.6%; Pred. No. 5e-240;

XX CC Matches 520; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRLRVLLKTTWLEVPETPTGLHRLSLSLCTWGYSSNTRFTITLNYKDLPTGDE 60

DB 1 mrlrvllkttwlevpeteptglhrlslslctwgyssntrftitlnykdlptgde 60

QY 61 ETLASYGIVSGDLICILQDDIPAPNIPSTSESHSLONNQPSTLATSSNQTSMQDEQP 120

DB 61 etlasygivsgdlilclqddipapnipsstsehslnqneqpslatssnqtsmqdeqp 120

QY 121 SDSFQGAAGSGVWDDSMGLPSONTEAEISIQDNAMHAECTGTGYPSEPMLCSESVGGQVP 180

DB 121 sdsfqqgaagsgvwnddsmglpsonteaeisiqdnahmaegtgtfypsepmLCseSVGGQVP 180

QY 181 HSLETLQASDCSDANDALIVLHLLMESGYIPQGTAEAKALSMPEKWLKSLGYKLYVMH 240

DB 181 hsletlqasdcsdandallivllhllmesgyipqgtaeakalsmpekwlksglyklyvmh 240

QY 241 PLCEGSSATLTCVPLGNLVNATLTKNNEIRSVKRLQLLPESFICKELGENVANIYKD 300

DB 241 plcegssatltcvplgnlvnattknneirsvkrlllpesfickelgenvanykd 300

QY 301 LQKSLRFLKDLVYPLLAFTQALNLPDVLGLVPLLEKLRLFRLLDVSLSLSAVCR 360

DB 301 lqksrlfkdqlyvpllaftqalnlpdvlgvlplleklrlfrlldvsvlsisavcr 360

QY 361 DLFTASNDPLLRFLYLRDFRONTVRVQDQDKELYKRHTQKESPKGRFVMLLPSSSTH 420

DB 361 dlftasndpllrfllylrdfrrontvrvqtdwkelykrhtqkespkgrfvmllpssth 420

QY 421 TTPFVNPPLHPPFPSSRLPPIGGEDQRPPLFYVGDPISSILIPGGETPSPQFPPLRP 480

DB 421 ttpfvpnlhprfssrlppiggedqrptlpyvgdpsissilipggetpsqfpplrp 480

QY 481 RFDVGLPGLPGLPILPGRGPNDRFPFRPSGRPTDGRLSFM 522

DB 481 rfdvglpplpplpgrgpnldrffprpsgrptdgrlsfm 522

RESULT 2

AAW68521

ID AAW68521 standard; Protein: 591 AA.

XX CC

XX CC AAW68521;

XX CC

XX CC 25-JAN-1999 (first entry)

XX CC

XX CC Human RIP-associated protein.

XX CC

XX CC Human; RIP-associated protein; RAP; primer; PCR; amplification; probe;

XX CC hybridisation; death domain; MORT domain; ICE-like family protease;

XX CC kinase; TRAF domain; inflammation; cell death; tumour; HIV; infection.

XX CC

OS Homo sapiens.

XX PN WO9841624-A1.

XX PD 24-SEP-1998.

XX PF 19-MAR-1998; 98WO-IL00125.

XX PR 19-MAR-1997; 97IL-0120485.

XX PA (YEDA) YEDA RES & DEV CO LTD.

XX PI Kovalenko A, Wallach D;

XX DR WPI; 1998-531565/45.

XX DR N-PSDB; AAV57200.

XX CC DNA encoding RIP-associated protein (RAP) - useful for, e.g.

XX CC treatment of tumour cells or HIV-infected cells

XX CC Claim 10; Fig 2; 65pp; English.

XX CC This sequence represent part of a human RIP-associated protein (RAP).

XX CC The coding sequence was isolated from a B-cell library by a yeast

XX CC 2-hybrid screen using the RIP protein devoid of its "death domain" as

XX CC a bait. The screen isolated a clone of about 1.9 kb. Primers were

XX CC generated based on the sequence and used to PCR amplify probes for

XX CC screening a colon and heart cDNA library. A further 300 bp of sequence

XX CC was determined, which was added to the 1.9 kb of sequence from the B-cell

XX CC library. The encoded protein does not contain a "death domain", MORT

XX CC MODULE, ICE-like family protease domain, kinase domain, nor TRAF-domains.

XX CC RAP was shown to bind only to RIP and not to TRADD, MORT-1, p55-R, p75-R

XX CC or MACH. The protein can be used to modulate or mediate RIP

XX CC modulated/mediated intracellular effects on the inflammation, cell death

XX CC or cell survival pathways in which RIP is involved, e.g. for treating

XX CC tumour cells or HIV-infected cells.

XX CC

XX CC Sequence 591 AA;

XX CC

QY 42 SNTRFTITLNYKDLPTGDEETLASYGIVSGDLICILQDDIPAPNIPSTSESHSLQNN 101

DB 113 sntrftitlnykdlptgdeetlasygivsgdlilclqddipapnipsstsehslnqnn 172

QY 102 EQPSLATSSNQTSMQDEQPSDFQGAAGSGVWDDSMGLPSONTEAEISIQDNAMHAEGT 161

DB 173 eq--latsnqtSMQDEQPSDFQGAAGSGVWDDSMGLPSONTEAEISIQDNAMHAEGT 230

QY 162 GFYSPPEMLCSESVGGQVPFHSLETLYQASDCSDANDALIVLHLLMESGYIPQGTAEAK 221

DB 231 gfypsepmLCseSVGGQVPFHSLETLYQASDCSDANDALIVLHLLMESGYIPQGTAEAK 290

QY 222 LSMPEKWLKSLGYKLYVMHPLCEGSSATLTCVPLGNLVNATLTKNNEIRSVKRLQLLP 281

DB 291 lsmpekwlksglyklyvmhplcegssatltcvplgnlvnattknneirsvkrlllp 350

QY 282 ESFICKELGENVANIYKDLQKLSRLFKDQVLYPLLAFTQALNLPDVLGLVPLLEKL 341

DB 351 esfickelgenvanykdlqksrlfkdkqlyvpllaftqalnlpdvlgvlpllekl 410

QY 342 RIFRLDVSRLSVLSAVCRDLFTASNDPLLRFLYLRDFRONTVRVQDQDKELYKRHT 401

DB 411 rifrlldvsvlsisavcrdlftasndpllrfllylrdfrrontvrvqtdwkelykrht 470

QY 402 QRKESPKGRFVMLLPSSSTHTTIPFVNPPLHPPFPSSRLPPIGGEDQRPPLFYVGDPI 461

DB 471 qrkespkgrfvmllpssththtippfvpnlhprfssrlppiggedqrptlpyvgdpl 530

QY 462 SSLIPGGETPSPQFPPLRPFRFPDRFPVGLPGLPILPGRGPNDRFPFRPSGRPTDGRLSF 521

XX CC

XX CC Query Match 91.1%; Score 2510; DB 19; Length 591;

XX CC Best Local Similarity 99.6%; Pred. No. 1e-218;

XX CC Matches 479; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

|||||
Db 531 sslipggetpsqplprfrdvpplbgpnpilpgrgpnndrfprpsrgrptdgrlsf 590
Qy 522 M 522
Db 591 m 591
RESULT 3
AAV83047
ID AAV83047 standard; Protein; 482 AA.
XX
AC AAV83047;
XX
DT 16-AUG-2000 (first entry)
XX
DE F-box protein FBP-7.
XX
KW F-box protein; FBP; diagnosis; treatment; screening; agonist;
KW antagonist; proliferative disorder; differentiative disorder;
KW breast cancer; prostate cancer; ovarian cancer; cancer;
KW small cell lung carcinoma; immune disorder; cardiovascular disorder;
KW inflammatory disorder; human.
XX
OS Homo sapiens.
XX
PN WO200012679-A1.
XX
PD 09-MAR-2000.
XX
PF 27-AUG-1999; 99WO-US19560.
XX
PR 28-AUG-1998; 98US-0098355.
PR 03-FEB-1999; 99US-0118568.
PR 15-MAR-1999; 99US-0124449.
XX
PA (UUNY) UNIV NEW YORK STATE.
XX
PI Chiaur DS, Pagano M, Latres E;
XX
DR WPI; 2000-256635/22.
DR N-PSDB; AAZ93356.
XX
PT Novel nucleic acid for screening compounds useful for treating
PT proliferative and differentiative disorders such as cancer and immune
PT disorders comprises sequences encoding ubiquitin ligases -
XX
PS Claim 10; Figure 10a; 245pp; English.
XX
CC Nucleic acids encoding substrate-targeting subunits of ubiquitin
CC ligases with F-box motifs (F-box proteins) are useful for diagnosis
CC of proliferative and differentiated related disorders by measuring
CC FBP gene expression. Cells expressing such proteins or
CC their fragments are useful for screening compounds. The compounds
CC are agonists or antagonists, which are useful for treating a
CC proliferative or differentiative disorder in a mammal such as
CC breast, ovarian and prostate cancer and small cell lung carcinoma
CC and also major opportunistic infections, immune disorders,
CC cardiovascular diseases and inflammatory disorders. FBP protein,
CC analogs, derivatives and their subsequences, anti-FBP antibodies
CC are also useful in diagnosis of the disorders.
XX
SQ Sequence 482 AA;
Query Match 90.4%; Score 2489; DB 21; Length 482;
Best Local Similarity 98.1%; Pred. No. 6.1e-217;
Matches 472; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Qy 42 SNRTFTITLNKDPITGDEETLASYGIVSGDLICLILLODDIPAPNPISSTDSHSSLNQ 101
Db 2 snrtftitlnykdpitgdeetlasygivsgdlclilhdhddipnpisdstshsslnq 61

Qy 102 EQPSLATSSNOTSMODEQSDSQGAAQSGVWDDSMGLPSQNFPAESIQDNAHMAEQT 161
Db 62 eqpslatssnqtsiqdeqpsdsfqgaaqsgvwnddsmglpsqnfpaesiqdnahmaegt 121
Qy 162 GFYSEPMLCSESVGQVPHSLSTLYQSADCSANDALIVLIHLLMLESYGYPQTEAKA 221
Db 122 gfysepllcsevesvegqphsletlyqsadcsandalivlihlmllesgyipgqteaka 181
Qy 222 LSMPEKWKLSGVYKLOYMHPLCEGSSATLTCVPLGNLIWVWATLKINNEIRSVKRLQLLP 281
Db 182 lsipekwklsykvylqymhhlcegsatltcvpignliwvwnatlkinneirsvkrlqlp 241
Qy 282 ESFTCKEKLGENVANIYKDLQKLSRFLKDLVYPLLAFTRAQNALNPVDFGLVVLPLEKL 341
Db 242 esfickeklgenvaniykdqlksrlfkdlvypyllafrqalnlpnvfglvvlplekl 301
Qy 342 RIFRLLDVRSVLSLAVCRDLFTASNDPLLMRFYLRLFRONTVRVQDTHKELYRKRI 401
Db 302 rifrlldvrsvislsavcrdlftasndpllmrfylrldrntvrvqtdwkelyrkrhi 361
Qy 402 QRKESPKGRFVMLPSSTHTIPFYPNPLHPRFPSSRLPPGIIGGEYDQRPTLPYVGDP 461
Db 362 qrkespkgrfvmlpssthtipfypnplhprfpssrllppgiiggeyqrptlpyvgdp 421
Qy 462 SSLIPGGETPSQFPLPRFRDVPGLPGPNPILPGRGPNDRPFRPSRGRPTDGRUSF 521
Db 422 sslipggetpsqlplprfrdvpglpgpnpilpgrgpnndrfprpsrgrptdgrisf 481
Qy 522 M 522
Db 482 m 482
RESULT 4
AAU32109
ID AAU32109 standard; Protein; 607 AA.
XX
AC AAU32109;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2600.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.
XX
PR 26-JAN-2001; 2001US-0770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 556; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for

CC Identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

XX SQ Sequence 607 AA;

Query Match 88.9%; Score 2448.5; DB 22; Length 607;
Best Local Similarity 90.8%; Pred. No. 4.1e-213;
Matches 481; Conservative 7; Mismatches 33; Indels 9; Gaps 5;

OY 1 MRURVRLKRTWPLEVPEPTGLHRLSHRLSLICTWGYSSNTRFTITLVNKDPLTGDE 60
DB 36 mrlrvrlkrtwplevpetegtlghrlshrlsllctwgyssntrftitlvnkdpdtgde 95
OY 61 ETLASYGVSGDLICILQDIPAPNIPSTDSEHSLQNNQPSLATSSNOTSMODEQP 120
DB 96 etlasygvsgdlilcldqipapnipstdsehslqnnqpslatssngtmsmdqep 155
OY 121 SDSFGQAAGSGVWDDSMGLPSONFEASIQDNAHMAECTGYPEPMCLSESVEGQVP 180
DB 156 sdsfggaagsgvwnddsmglpsnqfeasiqdnahmaectgyfypsepmclsesveggvp 215
OY 181 HSLFTLYQSADCSANDALIVLHLLMESGYIPOQTEAKALSMPEKWKLSGVYKLYQYMH 240
DB 216 hslftlyqsadcsandallivlhllmlesgyipqgteakalsmpekwklsgvyklqymh 275
OY 241 PLCGSSSATTCVPLGLNIVVNATLTKINNEIRSVKRLQLLPESFICKELGENVANLYKD 300
DB 276 plcgssatitcvplglnivvnatlkinneirsvkrqlllpessfickeligenvanlykd 335
OY 301 LOKLSRLFKDQVYPLLAFTROALNLPDVFGLVPLLEKLIRFLRDLVRSVLSLAVCR 360
DB 336 lqklsrlfkdqvypyllafrqalnlpdvglvplleklirflrldvrsvlsavcr 395
OY 361 DLFTASNDPLLRFLYLRDRFNDTVRVQDTEWELYKRHIQKESPKGRFVMLLPSSSTH 420
DB 396 dlftasndpllrfllylrdrfndtvrgqtdwkelykrhiqrkespkgrfvmllpssth 455
OY 421 TIPFPNPLHPRPSSRLPGIIGGYDORP-TLPVVGDPISSLIPCP-GETPS-QFPP 477
DB 456 tipfynplhprpssrlpgiiggydtpktntcfmldpisslipwvlgetpqsfp 515
OY 478 -----LRPRFPVGPLPGPNILPGRGPNDRFPFRSGRPTDGRLSFM 522
DB 516 tethalnpwppisdrnpqpcgaegppnrt-rfpplrqp-ggranlm 564

RESULT 5
AAB35160
ID AAB35160 standard; Protein; 443 AA.

AC AAB35160;

XX 09-APR-2001 (first entry)

DE Human Skp1-associated F-box protein-1 SAF-lalpa SEQ ID NO: 8.

XX Human; protein degradation; slah-mediated degradation protein; SMDP;
KW SCF-complex protein; SCP; slah-lalpa; slah-1 interacting protein; SIP;
KW Skp1-associated F-box protein; SAF-1; SAF-2; SAD; cancer; cell division;
KW Skp1-associated destruction-box protein; inflammatory disease.

XX

OS Homo sapiens.

XX WO200077207-A2.

XX 21-DEC-2000.

XX 09-JUN-2000; 2000WO-US15873.

XX 11-JUN-1999; 99US-0330517.

XX (BURN-) BURNHAM INST.

XX Reed JC, Matsuzawa S;

XX WPI; 2001-071273/08.

XX N-PSDB; AAC67284.

XX Slah-Mediated Degradation Protein, useful for drug screening, for
XX therapeutic applications and for functional genomics -

XX Claim 15; Page 103-104; 121pp; English.

XX The present invention provides the protein and coding sequences of
CC several slah-mediated degradation proteins and SCF-complex proteins.
CC These are designated Slah-lalpa, Slah-1 interacting protein (SIP), which
CC encodes two proteins due to alternative splicing (SIP-L and SIP-S),
CC Skp1-associated F-box protein-lalpa and beta and -2 (SAF-lalpa,
CC SAF-lbeta and SAF-2) and Skp1-associated destruction-box protein (SAD).
CC The proteins and their coding sequences are useful in the diagnosis and
CC treatment of cancers, disorders where too little cell division occurs
CC such as bone marrow aplasia, immunodeficiencies and inflammatory
CC diseases including sepsis, fibrosis, arthritis and graft versus host
CC disease.

XX SQ Sequence 443 AA;

Query Match 82.4%; Score 2270; DB 22; Length 443;
Best Local Similarity 99.8%; Pred. No. 4e-197;
Matches 430; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 92 DSEHSLQNNQPSLATSSNOTSMODEQPSDSFGQAAGSGVWDDSMGLPSONFEAS 151

DB 13 dsehslqnnqpslatssngtmsmdqpsdsfggaagsgvwnddsmglpsnqfeas 172

OY 152 QDNAHMAECTGYPEPMCLSESVEGQVPHSLFTLYQSADCSANDALIVLHLLMESG 211

DB 73 qdnahmaectgyfypsepmclsesveggvphsletlyqsadcsandalivlhllmesg 132

OY 212 YTPQGTAKALSMPEKWKLSGVYKLYQYMHPLCEGSSATTCVPLGLNIVVNATLKNNEI 271

DB 133 ytpqgteakalsmpekwklsgvyklqymhplcegsatitcvplglnivvnatlkinnei 192

OY 272 RSVKRLQLLPESFICKELGENVANLYKDQLSRFLKDLQVYPLLAFTROALNLPDVF 331

DB 193 rsvkrqlllpessfickeligenvanlykdqlsrflkdlqvlpyllafrqalnlpdvfg 252

OY 332 LVVLPLEKLIRFLRDLVRSVLSLAVCRDLFTASNDPLLRFLYLRDRFNDTVRVQD 391

DB 253 lvvlpieklirflrldvrsvlsavcrdlftasndpllrfllylrdrfndtvrvqdt 312

OY 392 WKELYKRHIQKESPKGRFVMLLPSSSTHTIPFPNPLHPRPSSRLPGIIGGYDOR 451

DB 313 wkelykrhiqrkespkgrfvmllpssthtipfynplhprpssrlpgiiggydqr 372

OY 452 PTLPYVGDPISSLIPGGETPSQFPPLRPRFDPVGPLPGPNILPGRGPNDRFPFR 511

DB 373 ptlpyvgdpisslipggetpsqfplprfdpvgplpgpnilpgrgpnldrfrpr 432

OY 512 GRPTDGRLSFM 522

DB 433 grptdgrlsfm 443

XX PD 06-JAN-2000.
XX XX
XX PF 25-JUN-1999; 99WO-US14484.
XX PR 26-JUN-1998; 98US-0090762.
XX PR 31-JUL-1998; 98US-0094983.
XX PR 01-OCT-1998; 98US-0102686.
XX PR 11-DEC-1998; 98US-0112129.
XX PA (INCY-) INCYTE PHARM INC.
XX XX
XX PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
XX PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
XX PI Bandman O;
XX XX
XX DR WPI: 2000-160673/14.
XX DR N-PSDB; AA298241.
XX XX
XX PT New human signal peptide-containing proteins useful in treatment,
XX PT prevention and diagnosis of e.g. cancer, inflammation and
XX PT cardiovascular disease
XX XX
XX PS Claim 1; Page 251; 327pp; English.
XX XX
XX CC AA298109 to AA298242 encode AAY87224 to AAY87357 which represent the
XX CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have
XX CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
XX CC neuroprotective, cardiovascular and antiasthmatic activities, and can
XX CC be used in gene therapy. HSPs can be used to treat or prevent disorders
XX CC associated with decreased activity or function of HSP. Antagonists of
XX CC HSP are used to treat or prevent disorders associated with increased
XX CC activity or function of HSP. Such diseases include cell proliferation
XX CC (including cancer), inflammation, cardiovascular, neurological,
XX CC reproductive or developmental disorders, (e.g. arteriosclerosis,
XX CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
XX CC asthma, Crohn's disease, microbial or other infections, congestive or
XX CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
XX CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP
XX CC nucleic acids can be used for the recombinant production of HSP, for
XX CC detecting HSP in standard hybridisation and amplification assays (for
XX CC diagnosis and monitoring), in gene therapy, as antisense,
XX CC triplex-forming or ribozyme therapeutics, for detecting related sequences
XX CC or genetic variations, and for chromosomal mapping. HSP are also used to
XX CC raise specific antibodies (Ab) and to screen for agonists and
XX CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
XX CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic
XX CC antagonists, in competitive drug screens, and for purification of HSP
XX CC from natural sources.
XX XX
XX SQ Sequence 113 AA;

Query Match 10.9%; Score 299; DB 21; Length 113;
Best Local Similarity 80.5%; Pred. No. 3.6e-19;
Matches 62; Conservative 4; Mismatches 7; Indels 4; Gaps 1;

Qy 309 KDQLVPLAFTR---QALNLPDVGVLVPLPLEKLRIFRLLDVRSVLSAVCRDLFT 364
| : : : : |
Dy 4 kpdvlysvipvtsifflalnlpdvfglvplplekkrifrlldvrsvlsavcrdlft 63
| : : : : |
Qy 365 ASNDPLLRFLYLRDR 381
| : : : : |
Dy 64 asndpllrfllylrdr 80
| : : : : |

RESULT 10
ABG06838
ID ABG06838 standard; Protein; 225 AA.
XX AC
XX AC ABG06838;
XX XX
XX DT 13-FEB-2002 (first entry)

XX XX Novel human diagnostic protein #6829.
XX DE
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX XX
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI
XX DR WPI: 2001-639362/73.
XX DR N-PSDB; AAS71025.
XX XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX PS Claim 20; SEQ ID No 37197; 103pp; English.
XX XX
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 225 AA;

Query Match 9.6%; Score 265.5; DB 22; Length 225;
Best Local Similarity 69.9%; Pred. No. 1.1e-15;
Matches 58; Conservative 5; Mismatches 19; Indels 1; Gaps 1;

Qy 329 VFGLVPLPLEKLRIFRLLDVRSVLSAVCRDLFTASNDPLLRFLYLRDRFTVTRVQ 388
| : : : : |
Dy 3 vfglvplplekkrifrlldvrsvlsavcrdlftasndpllrfllylrdrfrgd-frnd 61
| : : : : |
Qy 389 DTDWKELRYRKRHIQRKSPKGRF 411
| : : : : |
Dy 62 iyswttkdefyynlvdktearf 84
| : : : : |

RESULT 11
ABG18509
ID ABG18509 standard; Protein; 53 AA.
XX XX

ABG18509;
 18-FEB-2002 (first entry)
 Novel human diagnostic protein #18500.
 Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.
 Homo sapiens.
 WO200175067-A2.
 11-OCT-2001.
 30-MAR-2001; 2001WO-US08631.
 31-MAR-2000; 2000US-0540217.
 23-AUG-2000; 2000US-0649167.
 (HYSE-) HYSEQ INC.
 Drmanac RT, Liu C, Tang YT;
 WPI; 2001-639362/73.
 N-PSDB; AAS82696.
 New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity -
 Claim 20; SEQ ID No 48868; 103pp; English.
 The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. ABG00010-ABG30377 represent novel human
 diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 53 AA;
 Query Match 9.6%; Score 264.5; DB 22; Length 53;
 Best Local Similarity 56.5%; Pred. No. 1.5e-16;
 Matches 52; Conservative 0; Mismatches 1; Indels 39; Gaps 1;
 QY 431 PRPFSSRLPGIIGEGYDQRTPLPVYGDPISSLIIPGGETPSQFPPLRPDPVGLPG 490
 Db 1 prpfksrlpgilggydqr----- 21
 QY 491 PNPILPGRGGPNDRFPFRPSRGRPTDGRLSFM 522
 Db 22 pnpilpgrggpnldrfrpsrgrptdgrlsfm 53
 RESULT 13
 AAG29404
 ID AAG29404 standard; Protein: 475 AA.
 XX

AAU32108
 ID AAU32108 standard; Protein: 53 AA.
 XX
 AC AAU32108;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #2599.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20; Page 556; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SQ Sequence 53 AA;

Query Match 9.6%; Score 264.5; DB 22; Length 53;
 Best Local Similarity 56.5%; Pred. No. 1.5e-16;
 Matches 52; Conservative 0; Mismatches 1; Indels 39; Gaps 1;
 QY 431 PRPFSSRLPGIIGEGYDQRTPLPVYGDPISSLIIPGGETPSQFPPLRPDPVGLPG 490
 Db 1 prpfksrlpgilggydqr----- 21
 QY 491 PNPILPGRGGPNDRFPFRPSRGRPTDGRLSFM 522
 Db 22 pnpilpgrggpnldrfrpsrgrptdgrlsfm 53

RESULT 13
 AAG29404
 ID AAG29404 standard; Protein: 475 AA.
 XX

AC AAG29404;
XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 34981.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127452.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

```
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 240.5; DB 21; Length 475;
Best Local Similarity 24.5%; Pred. No. 6.6e-13;
Matches 116; Conservative 89; Mismatches 177; Indels 91; Gaps 20;

QY 1 MRLVRLLKRTWPEPTLGLHLSRLSLCTGWYSSNTRFTITINVKDPL--TG 58
DB 3 lrrhretklleladad-tlhdlrrlnptc-----pssvhlslnrkdelitps 53
QY 59 DEETLASYGIVSGDLICLIQLDDIPAPNIPSSPTDSEHSLQNNQPSLATSSNQTSMD 118
DB 54 pedtlrslgldsllyfsle-----agesnwkldrtsetvasqsesnqtsvhd- 102
QY 119 QPSDSFGQQAQGVVNDMSLGPONFAESTQDNHMAEG--TGFYSPSPM----- 169
DB 103 --sigf----aevdvypdqaknnpn-----tsved----pegdisgmegpepmdveqldm 147
QY 170 --LCSESVGEQVPHSLTET--LYSADCSANDALIVLIHLMLSCYI--POGTE----- 218
DB 148 elaaagskrlsepfllknilleksdgtseit-claisvhamlesgfvllnhgskfnfs 206
```

```
QY 219 AKALSMPEKWLKSGYVKLOYMHPLECGSSATLTCVPLGNLIYVNATL-----KINNEI-- 271
DB 207 keltvtsrytlpelikskdtnti---easvskfqnlgpvvvvygtvgssgrvhhmldk 263
QY 272 -RSVKRLQLLPESFICKELGENVANIYKDLQKLSRLFKDQLVYPLLAFTQALNLPDVF 330
DB 264 rrfvpvldvmdtsdsde---egsssiyrevfmfwmvkdrlvlpilgicdkaglep 320
QY 331 GLVPLPLELKLRIFRLLDVRSLSAVCRDLFTASNDPLLRFLYLRDRFRONTVRV--O 388
DB 321 clmrlptelklilellpgvsignmacvctemrylasndnlwkqcleevnnfvvteagd 380
QY 389 DTDWK-----ELYKRRIHQKESPKGRFVMLLPSSHTIFP 424
DB 381 svnwkarfatfwrqklaasdtfwrqnlgrnlistgrsgrfprlqgdpf 433

RESULT 14
AA83054
ID AA83054 standard; Peptide; 39 AA.
XX
AC AA83054;
XX
DT 16-AUG-2000 (first entry)
XX
DE F-box motif of FBP-6.
XX
KW F-box protein; FBP; diagnosis; treatment; screening; agonist;
KW antagonist; proliferative disorder; differentiative disorder;
KW breast cancer; prostate cancer; ovarian cancer; cancer;
KW small cell lung carcinoma; immune disorder; cardiovascular disorder;
KW inflammatory disorder; human.
XX
OS Homo sapiens.
XX
PN WO200012679-A1.
XX
PD 09-MAR-2000.
XX
PF 27-AUG-1999; 99WO-US19560.
XX
PR 28-AUG-1998; 98US-0098355.
PR 03-FEB-1999; 99US-0118568.
PR 15-MAR-1999; 99US-0124449.
XX
PA (UUNY ) UNIV NEW YORK STATE.
XX
PI Chiaur DS, Pagano M, Latres E;
XX
DR WPI; 2000-256635/22.
XX
PT Novel nucleic acid for screening compounds useful for treating
PT proliferative and differentiative disorders such as cancer and immune
PT disorders comprises sequences encoding ubiquitin ligases -
XX
PS Disclosure; Page 197; 245pp; English.
XX
CC Nucleic acids encoding substrate-targeting subunits of ubiquitin
CC ligases with F-box motifs (F-box proteins) are useful for diagnosis
CC of proliferative and differentiative related disorders by measuring
CC FBP gene expression. Cells expressing such proteins or
CC their fragments are useful for screening compounds. The compounds
CC are agonists or antagonists, which are useful for treating a
CC proliferative or differentiative disorder in a mammal such as
CC breast, ovarian and prostate cancer and small cell lung carcinoma
CC and also major opportunistic infections, immune disorders,
CC cardiovascular diseases and inflammatory disorders. FBP protein,
CC analogs, derivatives and their subsequences, anti-FBP antibodies
CC are also useful in diagnosis of the disorders.
XX
SQ Sequence 39 AA;
```

as
TTTTTTT T

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 11:28:11 ; Search time 15.29 seconds
(without alignments)
833.889 Million cell updates/sec

Title: US-09-927-458-2

Perfect score: 2754

Sequence: 1 MLRLVRLKRTWPLEVETE.....DRFPFRSRGRTDGRLSFM 522

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	189	6.9	38	4	US-09-172-841-47
2	164	6.0	38	4	US-09-172-841-49
3	107	3.9	78	1	US-08-487-359-5
4	107	3.9	78	1	US-08-222-798A-5
5	103.5	3.8	79	1	US-08-487-359-8
6	103.5	3.8	79	1	US-08-222-798A-8
7	103.5	3.8	580	3	US-08-906-865-1
8	103.5	3.8	888	2	US-08-861-464-6
9	103.5	3.8	888	2	US-08-396-001-6
10	103.5	3.8	888	4	US-08-323-433A-6
11	102.5	3.7	79	1	US-08-487-359-2
12	102.5	3.7	79	1	US-08-222-798A-2
13	101.5	3.7	1149	3	US-08-560-005-5
14	101.5	3.7	1149	4	US-08-418-540-5
15	101	3.7	560	2	US-08-808-931-18
16	101	3.7	560	3	US-08-808-323-18
17	101	3.7	560	3	US-09-050-603A-18
18	101	3.7	560	3	US-09-102-420B-18
19	101	3.7	560	4	US-09-497-698-18
20	100.5	3.6	851	1	US-08-369-796-2
21	100.5	3.6	851	2	US-08-852-091-2
22	100.5	3.6	851	2	US-08-820-754-2
23	100.5	3.6	851	3	US-08-956-652-2
24	100.5	3.6	851	3	US-08-956-869-2
25	100.5	3.6	851	3	US-09-012-710-2
26	100.5	3.6	851	3	US-08-948-547-2
27	100.5	3.6	851	4	US-09-087-465-4

Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 13, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 37, Appli
Sequence 54, Appli
Sequence 54, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 5, Appli
Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-172-841-47
; Sequence 47, Application US/09172841
; Patent No. 6232081
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeffrey W.
; APPLICANT: Ellledge, Stephen J.
; TITLE OF INVENTION: F-BOX PROTEINS AND GENES
; FILE REFERENCE: BCM-03510
; CURRENT APPLICATION NUMBER: US/09/172.841
; CURRENT FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 08/951.621
; EARLIER FILING DATE: 1997-10-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-172-841-47

Query Match 6.9%; Score 189; DB 4; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.9e-11;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 LPLEKLKRIFRLLDVRVLSLSAVCRDLFTASNDPLW 372
DB 1 LPLEKLKRIFRLLDVRVLSLSAVCRDLFTASNDPLW 38
|||||

RESULT 2

US-09-172-841-49
; Sequence 49, Application US/09172841
; Patent No. 6232081
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeffrey W.
; APPLICANT: Ellledge, Stephen J.
; TITLE OF INVENTION: F-BOX PROTEINS AND GENES
; FILE REFERENCE: BCM-03510
; CURRENT APPLICATION NUMBER: US/09/172.841
; CURRENT FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 08/951.621
; EARLIER FILING DATE: 1997-10-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Mus musculus

```
us-09-172-841-49
Query Match          6.0%; Score 164; DB 4; Length 38;
Best Local Similarity 86.8%; Pred. No. 4.8e-09;
Matches 33; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 335 LPELKLRIFRLLDVRSLVLSAVCRDLFTASNDPLW 372
      |||||
Db 1 LPELKLRIFRLLDVRSLVLSAVCHDLLIASNDPLW 38

RESULT 3
US-08-487-359-5
; Sequence 5, Application US/08487359
; Patent No. 5633229
; GENERAL INFORMATION:
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: LEHRER, ROBERT I.
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; STATE: CA
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,798A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0553.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-222-798A-5

Query Match          3.9%; Score 107; DB 1; Length 78;
Best Local Similarity 37.4%; Pred. No. 0.0049;
Matches 34; Conservative 1; Mismatches 26; Indels 30; Gaps 7;

QY 423 PFYPNPLHRP-FPSSRLPPGIIGGYDQRTPLPVGDPISLLIPGGETPSQFPPLRP 481
      |||||
Db 5 PNVGPRFPFPQFGPRFP-----PN-----FPGPRFPFPQFP--GPR 41

QY 482 FDPVGPLPG---PNPILPGRGPNDRFP 509
      |||||
Db 42 FPP--PFGPFPFPFPVFGPFPFPFP--PFRP 69

RESULT 5
US-08-487-359-8
; Sequence 8, Application US/08487359
; Patent No. 5633229
; GENERAL INFORMATION:
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: LEHRER, ROBERT I.
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; STATE: CA
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
US-08-222-798A-5

us-09-172-841-49
Query Match          6.0%; Score 164; DB 4; Length 38;
Best Local Similarity 86.8%; Pred. No. 4.8e-09;
Matches 33; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 335 LPELKLRIFRLLDVRSLVLSAVCRDLFTASNDPLW 372
      |||||
Db 1 LPELKLRIFRLLDVRSLVLSAVCHDLLIASNDPLW 38

RESULT 3
US-08-487-359-5
; Sequence 5, Application US/08487359
; Patent No. 5633229
; GENERAL INFORMATION:
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: LEHRER, ROBERT I.
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; STATE: CA
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,359
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/222,798
; FILING DATE: 05-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0553.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-487-359-5

Query Match          3.9%; Score 107; DB 1; Length 78;
Best Local Similarity 37.4%; Pred. No. 0.0049;
Matches 34; Conservative 1; Mismatches 26; Indels 30; Gaps 7;

QY 423 PFYPNPLHRP-FPSSRLPPGIIGGYDQRTPLPVGDPISLLIPGGETPSQFPPLRP 481
      |||||
Db 5 PNVGPRFPFPQFGPRFP-----PN-----FPGPRFPFPQFP--GPR 41

QY 482 FDPVGPLPG---PNPILPGRGPNDRFP 509
      |||||
Db 42 FPP--PFGPFPFPFPVFGPFPFPFP--PFRP 69

RESULT 4
US-08-222-798A-5
```

Query Match 3.88; Score 103.5; DB 3; Length 580;

STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER TYPE: Floppy disk
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/396,001
FILING DATE: 28-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6408A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 888 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PS-08-396-001-6

[illegible]

RESULT 10
US-09-323-433A-6
; Sequence 6, Application US/09323433A
; Patent No. 621812
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James J.

```

? APPLICANT: Cole, Francesca
? APPLICANT: Kennedy, Brian
? TITLE OF INVENTION: GENES DETERMINING CELLULAR SENSENCE IN
? TITLE OF INVENTION: YEAST
? FILE REFERENCE: 0050.1491-003
? CURRENT APPLICATION NUMBER: US/09/323.433A
? CURRENT FILING DATE: 1999-06-01
? PRIOR APPLICATION NUMBER: US 08/396,001
? PRIOR FILING DATE: 1995-02-28
? PRIOR APPLICATION NUMBER: PCT/US94/09351
? PRIOR FILING DATE: 1994-08-15
? PRIOR APPLICATION NUMBER: US 08/107,408
? PRIOR FILING DATE: 1993-08-16
? NUMBER OF SEQ ID NOS: 48
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 6
? LENGTH: 888
? TYPE: PRT
? ORGANISM: Saccharomyces cerevisiae
? US-09-323-433A-6

```

[illegible]

RESULT 11
US-08-487-359-2
; Sequence 2, Application US/08487359
; Patent No. 5633229
; GENERAL INFORMATION:
; APPLICANT: KOKRYAKOV, VLADIMIR N.
; APPLICANT: HARWIG, SYLVIA S.L.
; APPLICANT: LEHRER, ROBERT I.
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.

STATE: CA
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487.359
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/222.798
FILING DATE: 05-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2000-0553.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 79 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-487-359-2

Query Match 3.7%; Score 102.5; DB 1; Length 79;
Best Local Similarity 38.5%; Pred. No. 0.013;
Matches 35; Conservative 0; Mismatches 27; Indels 29; Gaps 7;

QY 423 PFYPNLHPRP-FPSSRLPPGIIGGYDQRTLPYVGDPISSLIPGPGTTPSQPPPLRPR 481
Db 5 PNVFGPRFPNPPGPRFP-----PN-----IPGPRFPNPPNFP--GPR 41

QY 482 FDP---VGPLGPNPILPGCGNDRPFRP 509
Db 42 FPPNFPGP-PFPPPIFGPWFPFP-PFRP 70

RESULT 12
US-08-222-798A-2
Sequence 2, Application US/08222798A
Patent No. 5804553
GENERAL INFORMATION:
APPLICANT: KORYAKOV, VLADIMIR N.
APPLICANT: HARMIG, SYLVIA S.L.
APPLICANT: LEHRER, ROBERT I.
TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W.
CITY: Washington, D.C.
STATE: CA
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,798A
FILING DATE: 05-APR-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2000-0553.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 79 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-222-798A-2

Query Match 3.7%; Score 102.5; DB 1; Length 79;
Best Local Similarity 38.5%; Pred. No. 0.013;
Matches 35; Conservative 0; Mismatches 27; Indels 29; Gaps 7;

QY 423 PFYPNLHPRP-FPSSRLPPGIIGGYDQRTLPYVGDPISSLIPGPGTTPSQPPPLRPR 481
Db 5 PNVFGPRFPNPPGPRFP-----PN-----IPGPRFPNPPNFP--GPR 41

QY 482 FDP---VGPLGPNPILPGCGNDRPFRP 509
Db 42 FPPNFPGP-PFPPPIFGPWFPFP-PFRP 70

RESULT 13
US-08-560-005-5
Sequence 5, Application US/08560005
Patent No. 6001354
GENERAL INFORMATION:
APPLICANT: Pot, David A.
APPLICANT: Williams, Lewis T.
APPLICANT: Jefferson, Anne Bennett
APPLICANT: Majerus, Philip W.
TITLE OF INVENTION: No.6001354el Grb2 Associating Protein and Nucleic
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/560.005
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0624000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 1..1149
OTHER INFORMATION: /note= "51c"

Matches 120; Conservative 59; Mismatches 179; Indels 202; Gaps 29;

```
QY 16 VPETE-----PTLGLRSH-LRLSLCTW---GYSSNTRFTITLNYKDPLTCDEETLASY 66
Db 10 IPOTQCMPLRSSGHRGNCIMLSIPCSLIGRGYYSHKRRMSMCSCTS-SGSKSAVREA 68
QY 67 GIVSGD---LICILODDIPAPNIPSTDSEHSSIQNNEQPSLATSSNOTSMODE----- 118
Db 69 GSGGAGGLLDCVIVGGGTGICIAQALCTKHS-----SSLSNPFIVTEAKDRVGNI 122
QY 119 -----QISDSFQCAQAQSG-----VWDDSMGLGPSQ- 144
Db 123 VTVEADGYIWECPNSFQPSDAVLTMAYDVGKDELVLGDPNAPRFLVNDKLRVPVPSL 182
QY 145 ----NFEAESIQDNAHMAEGT-GFYPSPEMLCSEVGOVPHSL-----ETLYQSADGSD 194
Db 183 TOLPFDLMTIFGKTRAAALGALGFRPSPPP-HEESVEHFVRNLDGEYFERLIEPF-CSG 240
QY 195 ANDALIVLIHLLMESGYIPQTEAKALSMPEKWKLSGVYKLOYMHPLECGSSATLTCVP 254
Db 241 -----VYAGDPAK-LSM--KAAFGKWKLEQ-----KGG- 267
QY 255 LGNLIVVNATLKINNEIRS---VKRLQLLPESFICKELGENVANIIYKDQLKLSRLFKDQ 311
Db 268 -----IIGGTLKAIQERGSNPKPPRQRLP-----KPK-GQTVGSFRK----- 304
QY 312 LVYPLLAFTQALNLPDVFGLVPLLEKLRIFRLLDVRSLSLSAVCRDLFTASNDPLL 371
Db 305 -----GLVMLPTAISARL-----GSRVKLS 324
QY 372 WRFLYLRFDRNTVRVQPTDWKELYRKRIQRKESPKGRFVMLLPSSHTTIFPYNPPLHP 431
Db 325 WTL-----SSTVKSNGEYSITY-----DTPDGLSVYRTRKSVVMTVPSYVASRL 369
QY 432 RPPFSR-----LPPGIIGEYDQRPPLPVGDPISS--LIPGGETPSQFPPLRPR 481
Db 370 RPLSDSAADSLSKFYYP-----VAVVSLSYPREAIRSECLINGELOFGQLHPRSOG 422
QY 482 FDPVGPLPGPNPILPGRGGP 501
Db 423 VETLGTIYS-SSLFPGRAPP 441
```

Search completed: June 6, 2002, 11:30:44
Job time: 153 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 11:30:26 ; Search time 13.44 Seconds
(without alignments)
1503.840 Million cell updates/sec

Title: US-09-927-458-2
Perfect score: 2754
Sequence: 1 MRLRVRLKRTWPLEVETE.....DRFPFRSGRPTDGRLSFM 522

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2748	99.8	522	1 FBX7_HUMAN	Q9y311 homo sapien
2	147	5.3	270	1 PSF1_DROME	Q9v637 drosophila
3	144	5.2	447	1 FBX9_HUMAN	Q9uk97 homo sapien
4	123	4.5	271	1 PSF1_HUMAN	Q92530 homo sapien
5	120	4.4	684	1 CC4_CANAL	P53699 candida alb
6	117.5	4.3	503	1 WAIP_HUMAN	O43516 homo sapien
7	117.5	4.3	1213	1 FMN_CHICK	Q05858 gallus gall
8	116.5	4.2	592	1 ODP2_DICDI	P36413 dictyostell
9	114	4.1	2426	1 SON_HUMAN	P18583 homo sapien
10	111	4.0	1324	1 IRS2_HUMAN	Q9y4h2 homo sapien
11	110.5	4.0	1856	1 GBF1_CRIGR	Q9rl87 cricetus
12	110	4.0	424	1 S3B4_HUMAN	Q15427 homo sapien
13	110	4.0	1468	1 FMN1_MOUSE	Q05860 mus musculu
14	109.5	4.0	772	1 NFL1_HUMAN	Q14494 h nuclear f
15	109	4.0	741	1 NFL1_MOUSE	Q61985 mus musculu
16	109	4.0	1206	1 FM14_MOUSE	Q05859 mus musculu
17	109	4.0	3703	1 ABF1_HUMAN	Q15911 homo sapien
18	107.5	3.9	2404	1 SON_MOUSE	Q9qx47 mus musculu
19	107	3.9	505	1 ANXB_HUMAN	P50995 homo sapien
20	106.5	3.9	212	1 PF11_PIG	P51524 sus scrofa
21	106	3.8	1324	1 SYJ1_BOVIN	O18964 bos taurus
22	105.5	3.8	897	1 BGAL_CLOAB	P24131 clostridium
23	105	3.8	351	1 CBFL1_YEAST	P17106 saccharomyc
24	105	3.8	502	1 WASP_HUMAN	P42768 homo sapien
25	104.5	3.8	849	1 RSG2_HUMAN	Q15283 homo sapien
26	103.5	3.8	580	1 SYN3_HUMAN	O14994 homo sapien
27	103.5	3.8	888	1 YGB4_YEAST	P25339 saccharomyc
28	103	3.7	1157	1 SRM4_HUMAN	Q95104 homo sapien
29	102.5	3.7	228	1 PF12_PIG	P51525 sus scrofa
30	102.5	3.7	962	1 Y4E_SCHPO	O09731 schizosacch
31	102.5	3.7	1078	1 S24A_HUMAN	Q09486 homo sapien
32	102	3.7	159	1 RUL1C_HUMAN	P09234 homo sapien
33	102	3.7	159	1 RUL1C_MOUSE	Q62241 mus musculu

RESULT 1
FBX7_HUMAN

ID	FBX7_HUMAN	STANDARD	PRT	522 AA
AC	Q9Y311; Q9UKT2; Q9UF21;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	F-box only protein 7.			
GN	FBX07 OR FBX7			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE OF 41-522 FROM N.A.			
RX	MEDLINE=20003060; PubMed=10531035;			
RA	Cenciarelli C., Chlaour D.S., Guardavaccaro D., Parks W., Vida M., Pagano M.;			
RT	"Identification of a family of human F-box proteins."			
RL	Curr. Biol. 9:1177-1179(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20395565; PubMed=10945468;			
RA	Llyin G.P., Rialland M., Pigeon C., Guquen-Guillouzo C.;			
RT	"cDNA cloning and expression analysis of new members of the mammalian F-box protein family."			
RL	Genomics 67:40-47(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M., Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A., Baguley S., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P., Blakesley C., Bridgeman A.M., Buck D., Burgess J., Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M., Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N., Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M., Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E., Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J., Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M., Matthews L., Mccann O.T., Mcclay J., McLaren S., Mcmurray A.A., Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D., Williams L., Williams S., Bentley D.R., Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,			

Q15569 homo sapien
Q03387 triticum ae
O61329 mus musculu
O9fpq6 chlamydomon
P05143 mus musculu
O14924 homo sapien
Q9m330 arabidopsis
P39679 kluyveromyc
P26630 homo sapien
Q9r017 mus musculu
P18835 caenorhabdi
P26307 zea mays (m

RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L., Wang Q., Wang Z., White J., Willingham D., Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing J., Scheet P., Walker C., Wansley A., Wohlmann P., Pepin K., Nelson J., Korf I., Bedell J.A., Hillier L., Kurahashi H., Safta S., Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Safta S., Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L., Tiliakou H., Wright H., "The DNA sequence of human chromosome 22.;" Nature 402:489-495(1999).

CC -I- FUNCTION: probably recognizes and binds to some phosphorylated proteins and promotes their ubiquitination and degradation.

CC -I- SUBUNIT: PART OF A SCF (SKP1-CULLIN-F-BOX) PROTEIN LIGASE COMPLEX (BY SIMILARITY).

CC -I- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF129537; AAF04471.1; -

DR EMBL; AF233225; AAF67155.1; -

DR EMBL; AL050254; CAB43356.1; -

DR EMBL; 271183; CAB63143.1; -

DR MIM; 605648; -

DR InterPro; IPR001810; F-box.

DR Pfam; PF00646; F-box; 1.

DR SMART; SM00256; FBOX; 1.

DR PROSITE; PS50181; FBOX; 1.

KW Ubiquitin conjugation.

FT DOMAIN 329..375

FT CONFLICT 41..79 S -> M (IN REF. 1).

FT CONFLICT 79..79 Q -> H (IN REF. 1).

FT CONFLICT 84..84 A -> P (IN REF. 1).

FT CONFLICT 115..115 M -> I (IN REF. 1).

FT CONFLICT 169..169 M -> L (IN REF. 1).

FT CONFLICT 224..224 M -> L (IN REF. 1).

FT CONFLICT 241..241 P -> H (IN REF. 1).

FT CONFLICT 328..328 D -> N (IN REF. 1).

FT CONFLICT 413..413 M -> L (IN REF. 1).

FT CONFLICT 475..475 F -> L (IN REF. 1).

SQ SEQUENCE 522 AA; 58502 MW; C4E5E70A0747287A CRC64;

Query Match

Best Local Similarity 99.8%; Score 2748; DB 1; Length 522;

Matches 521; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRLVRLKRTWPLEVPETPTLGLHLSRLSLCTWGYSSNTFTITLNYKDLTGDE 60

DB 1 MRLVRLKRTWPLEVPETPTLGLHLSRLSLCTWGYSSNTFTITLNYKDLTGDE 60

QY 61 ETLSYGVSGDLICLLIQLDDIPAPNIPSTDSHSSLNQNEQPSLATSSNQTSMQDEP 120

DB 61 ETLSYGVSGDLICLLIQLDDIPAPNIPSTDSHSSLNQNEQPSLATSSNQTSMQDEP 120

QY 121 SDSFGQAAQSGWVNDSDMLGSPQNFEAESIQDNAHMAEGTGYFSEPMCLSESVEGVQVP 180

DB 121 SDSFGQAAQSGWVNDSDMLGSPQNFEAESIQDNAHMAEGTGYFSEPMCLSESVEGVQVP 180

QY 181 HSLEYLYQSADCSANDALIVLIHLLMESGYIPQGTAKALSMPEKWLKSLGVYKLYMH 240

DB 181 HSLEYLYQSADCSANDALIVLIHLLMESGYIPQGTAKALSMPEKWLKSLGVYKLYMH 240

QY 241 PLCEGSSATLTCVPLGNLIVVNTATLKINNEIRSVKRLQLLPESFICKELGENVANYKD 300

DB 241 PLCEGSSATLTCVPLGNLIVVNTATLKINNEIRSVKRLQLLPESFICKELGENVANYKD 300

QY 301 LQKLSRLFKDQLVYPLLAFTQALNLPDVFGLVPLPLELKLRIPLRLDVRSLVLSAVCR 360

DB 301 LQKLSRLFKDQLVYPLLAFTQALNLPDVFGLVPLPLELKLRIPLRLDVRSLVLSAVCR 360

QY 361 DLTASNDPLLRFLYLRDRFDNTVRVQDQTDWKLRYKRIQRKESPKGRFVMLLPSTH 420

DB 361 DLTASNDPLLRFLYLRDRFDNTVRVQDQTDWKLRYKRIQRKESPKGRFVMLLPSTH 420

QY 421 TIFYPNPLHPRFPSPSRLLPGGIIGGYDORPTLPVYGDPISSILIPGPGTGFPPPLRP 480

DB 421 TIFYPNPLHPRFPSPSRLLPGGIIGGYDORPTLPVYGDPISSILIPGPGTGFPPPLRP 480

QY 481 RFDVPGPLGPNILPGRGGPNDRFPFRPSRGRPTDGRLSFM 522

DB 481 RFDVPGPLGPNILPGRGGPNDRFPFRPSRGRPTDGRLSFM 522

RESULT 2

ID PSF1_DROME STANDARD; PRT; 270 AA.

AC Q9V637;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative proteasome inhibitor.

GN CG8379

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscophila;

OC Phyllophaga; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burkis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Flier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RESULT 4
 PSF1_HUMAN STANDARD; PRT; 271 AA.
 AC Q92530; Q9H411;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE proteasome inhibitor PI31 subunit (hPI31).
 GN PSMF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20309816; PubMed=10764772;
 RA McCutchen-Maloney S.L., Matsuda K., Shimbara N., Binns D.D.,
 RA Tanaka K., Staughter C.A., Demartino G.N.;
 RA "cDNA cloning, expression, and functional characterization of PI31, a
 RT proline-rich inhibitor of the proteasome.";
 RL J. Biol. Chem. 275:18557-18565(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.B.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhali P., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levassthalo M.H., Leversham M., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sulston J.E.,
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaadin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 CC -!- FUNCTION: Plays an important role in control of proteasome
 CC function. Inhibits the hydrolysis of protein and peptide
 CC substrates by the 20S proteasome. Also inhibits the activation of
 CC the proteasome by the proteasome regulatory proteins PA700 and
 CC PA28.
 CC -!- SUBUNIT: Monomer (Probable).
 CC -!- SIMILARITY: BELONGS TO THE PROTEASOME INHIBITOR PI31 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D88378; BAAL3603.1;
 CC EMBL: AL031665; CAC10383.1;
 CC Proteasome. 154 271 PRO-RICH.
 KW DOMAIN 36 36 C -> F (IN REF. 2).
 FT CONFLICT 36 36
 FT SEQUENCE 271 AA; 29772 MW; 40829596C1D665EE CRC64;
 SQ

Query Match 4.5%; Score 123; DB 1; Length 271;
 Best Local Similarity 23.0%; Pred. No. 0.15;
 Matches 81; Conservative 41; Mismatches 92; Indels 138; Gaps 22;
 QY 183 LETYQSA-----DCSDANDALIVLIHLLMLESYI-----PGTEAKALSMPEKWKLS 231
 DB 4 LEVLFAASAPAITC--RODALVCFLHVEVTHGYCGLVGQDGPDPNDKKSELLPAGMNN 61
 QY 232 -GVVKLOYMHPCLCEGSSATLTCVPLGNLIVVNATUKIN---NEIRSVKRLQLLPESFICK 287
 DB 62 KDLVIVLREYK--DGRKLLV-----KAITVSSMLNVLVSGQVADLTNLDDYIDA 114
 QY 288 EKLGNVANIYKDLQKL--SRLFKQDLVYVPLLAFTRAQLNLDPVGLVGLVLEKLRIERL 346
 DB 115 EHLGD-FHRTYKNSELSRI-----VSGII----- 139
 QY 347 LDVSVLSLAVCRDLFTASNDPLLRFLYLDRFDNTRVQDDTDWKELYRKRHIQRKES 406
 DB 140 ----- 153
 QY 407 PKGREVLLPSSTHTI-PFYPNPLHPRFPSSRLPPG-----IIGGEYDQRPTLPYVG 458
 DB 154 PHREPP--PATAREVDPLRIPPHPH--TSRQPPWCDPLGPFVVGGE-DLDPEPRRG 206
 QY 459 ----DPISS-----LIPGGETPSQFPP--LRP--RFDVVGPL-----PGNP 493
 DB 207 GMIVDPLRSRGPFRALIDPSSGLNRLPPGAVPGARFDPFGPTGTSPFPNP 258
 RESULT 5
 CC4_CANAL STANDARD; PRT; 684 AA.
 AC P53699;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cell division control protein 4.
 GN CDC4.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SGV126;
 RA Shieh J.C., White A.M., Rosamond J.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR INITIATION OF DNA
 CC REPLICATION AND SEPARATION OF THE SPINDLE POLE BODIES TO FORM THE
 CC POLES OF THE MITOTIC SPINDLE. IT ALSO PLAYS A ROLE IN BUD
 CC DEVELOPMENT, FUSION OF ZYGOTIC NUCLEI AFTER CONJUGATION AND
 CC VARIOUS ASPECTS OF SPOULGATION. REQUIRED FOR HTAI-HTB1 LOCUS
 CC TRANSCRIPTION ACTIVATION (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 1 F-BOX DOMAIN.
 CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X96763; CAA65538.1;
 CC InterPro: IPR001810; F-box.
 CC InterPro: IPR001680; WD40.
 CC Pfam: PF00646; F-box; 1.
 CC Pfam: PF00400; WD40; 7.
 CC PRINTS: PR00320; GPROTEINBRPT.
 CC SMART: SM00256; FBOX; 1.
 CC SMART: SM00320; WD40; 7.

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEGHORN; TISSUE=Embryo;
 RX MEDLINE=92112031; PubMed=1730407;
 RA Trump A., Blundell P.A., de la Pompe J.L., Zeller R.;
 RT "The chicken limb deformity gene encodes nuclear proteins expressed
 in specific cell types during morphogenesis.";
 RL Genes Dev. 6:14-28(1992).
 CC -|- FUNCTION: IS IMPORTANT FOR MORPHOGENESIS OF LIMB AND KIDNEY AND
 MAY BE INVOLVED IN DETERMINING DORSOVENTRAL NEURAL TUBE POLARITY
 AND MOTOR NEURON INDUCTION. IT MAY ALSO HAVE A FUNCTION IN
 DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING SPECIFIC
 DIFFERENTIATED STATES.
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE
 PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION
 IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE
 TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
 CC -|- TISSUE SPECIFICITY: PRESENT IN THE ADULT BRAIN, KIDNEY,
 BRAIN, HEART AND INTESTINE AND THROUGHOUT THE EMBRYO.
 CC -|- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN IS
 EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
 COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
 KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
 THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
 CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
 CC -|- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. CAPPUCCINO
 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: X62681; CAA44555.1;
 DR InterPro: IPR003104; FH2.
 DR InterPro: IPR001265; Formin.
 DR Pfam: PF02181; FH2; 1.
 DR PRINTS: PR00828; FORMIN.
 DR SMART: SM00498; FH2; 1.
 KW Nuclear protein; Developmental protein; Coiled coil;
 KW Alternative splicing.
 FT DOMAIN 428 450 COILED COIL (POTENTIAL).
 FT DOMAIN 503 572 COILED COIL (POTENTIAL).
 FT DOMAIN 652 751 FH1 (PRO-RICH).
 FT DOMAIN 766 1171 FH2.
 FT DOMAIN 1050 1125 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1213 AA; 135240 MW; ADE3EF03BF9D862 CRC64;
 Query Match 4.3%; Score 117.5; DB 1; Length 1213;
 Best Local Similarity 19.4%; Pred. No. 2.7;
 Matches 105; Conservative 55; Mismatches 204; Indels 177; Gaps 19;
 QY 57 TGEETLASGVISGDLICLIQDDIPAPNIPSTDSHSSLQNNQPSLATSSN-----111
 DB 297 TEGDETITEKPRENDLALLKQPVKKSNITSLGLTKKSKSPKASPTFLEQLSHLLNI 356
 QY 112 QTSWQBDQPSDSFQGAQAQGVNDDSLGSPQNFQAEISTQDNAHMAEGTGFYFSEPMLC 171
 DB 357 DVSKNDERTQDSAGRGE-----TEDSDGEPENKASGQ-----TEPLFP 395
 QY 172 SESVEGVQVPSHLETLYQSADCSANDALIVLIHL-----MLESYIPQGTQTEA-KALSMPE 226
 DB 396 SEETKS-----SPAEGALDVFKALFTRPKPKETTADPSELEAIKRWNE 440

QY 227 KWLKSGYVKLYQHPLCEGSSATLTCVPL-----GNLIYVNATLKINNEIRSVKRLQL 279
 DB 441 KESLKAVFERSKSKPGDPSDKSPDLSPSEQDDKTQGLQTVWPPKANHE-----491
 QY 280 LPESFICKELGENVANIYKDLQKLSRLFKDLQVYPLLAFTQALNLPDVGVLVLEL 339
 DB 492 -----EVKVGKLYTEAEYQAAILHLKREHKE-----IETLSQF 526
 QY 340 KLIRFLDLVRSVLSL-----SAVCRDLFTASND-----PILLWRFY 376
 DB 527 ELRVFHRTGEHAVSTAQLEETIAHLKNELDNKLNRNEEARDIGVSTEDDNLPTKYNVC 586
 QY 377 LRDRFRDNTVQDDTWKELYKRHIQKESKPGREVMLLPSSHTI-----422
 DB 587 IQDTRETFIKPSEENRAVKNQIVPKK-----LNISLTHSISTQGENKDSYDVPS 638
 QY 423 -----PFYVNPPLHPRFPSSRLPPGLIGGEYDQRTPLVVGDPISSLIP--466
 DB 639 SESVLSQCPKMLPSPSPPPPPPPPP-----PP-----PPPFSDSSLGLVPPP 684
 QY 467 -----GPGETPSQF-----PPLRPRF-----DPVGLPGNPILPGRGGPNDRFPFRSGR 513
 DB 685 PPLTGTPTSVTPHFAFGPPLPQLSEGCRDFAAPAPPAPPLPGLGPPVP--PPLPGSGL 742
 QY 514 P 514
 DB 743 P 743
 RESULT 8
 ODP2_DICDI
 ID ODP2_DICDI STANDARD; PRT; 592 AA.
 AC P36413;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
 DE complex, mitochondrial precursor (EC 2.3.1.12) (E2) (PDC-E2)
 DE (Fragment).
 GN DLAA.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
 RN NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX2;
 RA Mueller-Taubenberger A.;
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
 CC -|- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
 CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
 COPIES OF THREE ENZYMIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
 DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPONAMIDE DEHYDROGENASE
 (E3).
 CC -|- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide -> CoA + S-
 acetyldihydrolipoamide.
 CC -|- COFACTOR: THE E2 COMPONENT CONTAINS TWO COVALENTLY-BOUND LIPOYL
 COFACTOR.
 CC -|- SUBUNIT: 20 TO 30 ALPHA(2)-BETA(2) TETRAMERS OF E1 + 6 HOMODIMERS
 OF E3 + 60 COPIES OF E2.
 CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -|- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
 CC -|- SIMILARITY: CONTAINS 2 LIPOYL-BINDING DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: U06634; AAA16511.1; -


```
DR SMART: SM00222; Sec7; 1.
KW PROSITE: PS0190; SEC7; 1.
DR Guanine-nucleotide releasing factor.
FT DOMAIN 694
SQ SEQUENCE 1856 AA; 206190 MW; 46E8962C0E08EDBF CRC64;

Query Match 4.0%; Score 110.5; DB 1; Length 1856;
Best Local Similarity 20.3%; Pred. No. 14;
Matches 149; Conservative 74; Mismatches 207; Indels 303; Gaps 40;

QY 1 MRLVRLLKRWLEVPETPTGLGHRSHRLSLCTGWYSSNTRFTILNYKDPDPL----56
Db 1202 LRLAIRLLRR-----BEISQVLLSLRILLMKPVSLSRVSHQVAYGLHELLKTN 1252
QY 57 ----TGDE-----ETLASYGIVSGDLICLLIQQDIPAPNIPSPSTDSEHSSIQNNE--102
Db 1253 ANIHSGDDWATLTLLECIGS-GVKPPDQATARAD--APDAGASDSELPSYHONDVS 1309
QY 103 -----QPSLATS 109
Db 1310 LDRGYTSDSEVYTDHGRPGKIHRSATDADVMNSGVLVVGKDDIDNSKAGAGLSRPSPL 1369
QY 110 SNOTSMQ---DEQP-----SDSFQ-----CQAAQSGV 133
Db 1370 VNOYSLTVGLDGLPHDPTKSLKCVESLSFTVRDAAHITPDNFCVKTIRIFVEASLNGG 1429
QY 134 WNDSDMLGPGSQNEFAEISQDHAHMAECT-----GFYSPSEMLCSESVGEQV 179
Db 1430 CKSQDRKRGKSHKYSKGRNFKPKGKSVLRPRPTSSOCHTRGHSDE-----EEDG-V 1483
QY 180 PHSLET--LYOSADCSNDALIVLIHLMESGYIPQGTAKALSMPEKW-----228
Db 1484 PASVHTVSLQVSQD-----LLDLMHTL-----HTRAASITYSSWAEQORHLES 1525
QY 229 ---KLSGVYKLYMH---PLCEGSSATLTCVPLGNLIVNATLKINNEIRSVKRLQLLPE 282
Db 1526 GGRKIEADSTLWAHCWCPQLQG-IACLC-----DARRQVRMQL--1565
QY 283 SFTCKEKELGENVANIYKDLQKLSRL-----FKDQLVYPLLAFTROALNLPDVFGLVLPLE 338
Db 1566 TYLQR-----ALLVHDLQKDALEWESCFNKVLPRLTKLENISPADVGWME---E 1614
QY 339 LKLRIFLLDVRSVLSAVCRDLFTASNDPLLRFLYLRDFRD-----382
Db 1615 TRMRASLLSKVFLQHLS-----PLLSLSTFAALW--LTILDFMDKYNHAGSDLLSEAP 1668
QY 383 ----NTVRVQDT-----DWK-----ELYRKRHIQ--RKE 405
Db 1669 ESLKNLLVMDTAEIFHSADARGSPSALWEITWERTIDCFPLHRLDELFKQTVIQDPMPT 1728
QY 406 SPKGRFVMLLPSSHTTIPFYVNPPLHPRFPSSRLPPGIIGGEVDQ-----PT 453
Db 1729 EPHSQNAL---ASTHLTPAAGDPGH---LPSPEITPSEV--GACDSEKPEGTRATSSSPG 1780
QY 454 LPYVGDPDPISSLIPCP-GETSQSPP--LRPRFD-----VGPLPCP---NPLPGRGSPND 503
Db 1781 SPVASSP-SRLSPSPGPPPLAQPLLIQLPLTSLQVGVPPMALPIILNPALEATSP--1837
QY 504 RFPFRSGRGRPTD 516
Db 1838 -VPLL-STPRPTD 1848

RESULT 12
S3B4_HUMAN
ID S3B4_HUMAN STANDARD; PRT; 424 AA.
AC Q15427;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor 3B subunit 4 (Spliceosome associated protein 49) (SAP
DE 49) (SF3b50) (Pre-mRNA splicing factor SF3b 49 kDa subunit).
```

GN SF3B4 OR SAP49.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95047348; PubMed=7958871;
RA Champion-Arnaud P., Reed R.;
RT "The prespliceosome components SAP 49 and SAP 145 interact in a
RT complex implicated in tethering U2 snRNP to the branch site.";
RL Genes Dev. 8:1974-1983(1994).
RN [2]
RX CHARACTERIZATION OF THE SPLICEOSOME.
RP MEDLINE=20337962; PubMed=10882114;
RA Das R., Zhou Z., Reed R.;
RT "Functional association of U2 snRNP with the ATP-independent
RT spliceosomal complex E.";
RL Mol. Cell 5:779-787(2000).
CC -!- FUNCTION: SUBUNIT OF THE SPLICING FACTOR SF3B REQUIRED FOR 'A'
CC COMPLEX ASSEMBLY FORMED BY THE STABLE BINDING OF U2 SNRNP TO THE
CC BRANCHPOINT SEQUENCE (BPS) IN PRE-MRNA. SEQUENCE INDEPENDENT
CC BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS
CC ESSENTIAL. IT MAY ANCHOR U2 SNRNP TO THE PRE-MRNA. MAY ALSO BE
CC INVOLVED IN THE ASSEMBLY OF THE 'E' COMPLEX. SF3B4 HAS BEEN
CC FOUND IN COMPLEX 'B' AND 'C' AS WELL. BELONGS ALSO TO THE SPLICING
CC OF RARE CLASS OF NUCLEAR PRE-MRNA INTRON.
CC -!- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3B WHICH IS COMPOSED OF
CC FOUR SUBUNITS: SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145,
CC SF3B1/SAP155. SF3B ASSOCIATES WITH THE SPLICING FACTOR SF3A*AND A
CC 12S RNA UNIT TO FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEINS
CC COMPLEX (U2 SNRNP). SF3B4 INTERACTS DIRECTLY WITH SF3B2.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: STRONG, TO C.ELEGANS C08B11.5 AND, PARTIAL, TO
CC YEAST HSH49.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L35013; AAA60300.1; .
CC DR HSSP; P19339; 2SXL.
CC DR MIM; 605593; .
CC DR InterPro: IPR000504; RRM.
CC DR Pfam; PF00076; rrm; 2.
CC DR SMART; SM00360; RRM; 2.
CC DR PROSITE; PS50102; RRM; 2.
CC DR PROSITE; PS00030; RRM_RNP_1; 1.
CC KW Spliceosome; mRNA processing; mRNA splicing; Nuclear protein;
CC RNA-binding; Repeat.
CC FT DOMAIN 13 91 RNA-BINDING (RRM) 1.
CC FT DOMAIN 100 179 RNA-BINDING (RRM) 2.
CC FT DOMAIN 215 218 POLY-PRO.
CC FT DOMAIN 262 268 POLY-PRO.
CC SQ SEQUENCE 424 AA; 44385 MW; 212472A25D3FF002 CRC64;

Query Match 4.0%; Score 110; DB 1; Length 424;
Best Local Similarity 32.0%; Pred. No. 2.1;
Matches 39; Conservative 9; Mismatches 32; Indels 42; Gaps 9;
QY 430 HPRFPSSRLP-PGI-----IG-----GEYDQRTLPYVGDPISSLIPGP 469
Db 292 HPHPFPFGMPFGMSQMLAHGHGPHAGPGSGGQPPPPPPG-----MHPG 345
QY 470 ETSQSFPLPRPF-DPV---GPLP-----GPNILPGRG--GPNDRPFRPSRG-----R 513

Db 346 PPMGMPGPPGPGSPMGHPGMPHGMGPPPLMPHGHGTGPRPPPPYGYQGRGLPPPR 405
QY 514 PT 515
Db 406 PT 407

RESULT 13
ID FMNL_MOUSE STANDARD; PRT; 1468 AA.
AC Q05860;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Formin 1 isoforms I/II/III (limb deformity protein).
GN FMN OR ID.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney, and Testis;
RX MEDLINE=90363291; PubMed=2392150;
RA Woychik R.P., Maas R.L., Zeller R., Vogt T.F., Leder P.;
RT "Formins": proteins deduced from the alternative transcripts of the
RT limb deformity gene."
RL Nature 346:850-853(1990).
RN [2]
RP ALTERNATIVE SPLICING.
RX MEDLINE=97224459; PubMed=9119367;
RA Wang C.C., Chan D.C., Leder P.;
RT "The mouse formin (Fmn) gene: genomic structure, novel exons, and
RT genetic mapping."
RL Genomics 39:303-311(1997).
RN [3]
RP PHOSPHORYLATION.
RX MEDLINE=93296176; PubMed=8516300;
RA Vogt T.F., Jackson-Grusby L., Rush J., Leder P.;
RT "Formins: phosphoprotein isoforms encoded by the mouse limb deformity
RT locus."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5554-5558(1993).
CC -1- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
CC AND MAY HAVE A FUNCTION IN DIFFERENTIATING CELLS OR MAY BE
CC INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; IA (SHOWN HERE), IB,
CC II, III AND IV (AC Q05859); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC A VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
CC DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
CC -1- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
CC LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
CC GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
CC -1- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
CC IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
CC COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
CC KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
CC THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
CC -1- PTM: PHOSPHORYLATED ON SERINE AND POSSIBLY THREONINE RESIDUES.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FHL) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. CAPPUCCINO
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53599; CAA37668.1; -

DR PIR; S11515; S11515.
DR MGD; MGI:101815; Fmn.
DR InterPro: IPR003104; FH2.
DR InterPro: IPR001265; Formin.
DR Pfam: PF02181; FH2; 1.
DR PRINTS; PR00828; FORMIN.
DR SMART; SM00498; FH2; 1.
KW Nuclear protein; Developmental protein; Alternative splicing;
KW Phosphorylation; Coiled coil.
FT DOMAIN 723 792 COILED COIL (POTENTIAL).
FT DOMAIN 870 970 FHL (PRO-RICH).
FT DOMAIN 985 1426 FH2.
FT DOMAIN 1305 1378 COILED COIL (POTENTIAL).
FT DOMAIN 198 203 POLY-SER.
FT DOMAIN 861 864 POLY-SER.
FT DOMAIN 885 892 POLY-PRO.
FT DOMAIN 911 925 POLY-PRO.
FT DOMAIN 929 940 POLY-PRO.
FT DOMAIN 951 962 POLY-PRO.
FT DOMAIN 966 970 POLY-PRO.
FT DOMAIN 977 981 POLY-SER.
FT VARSPPLIC 1252 1287 MISSING (IN ISOFORM IB).
FT VARSPPLIC 625 722 MISSING (IN ISOFORM II).
FT VARSPPLIC 626 627 IA -> SV (IN ISOFORM III).
FT VARSPPLIC 628 1468 MISSING (IN ISOFORM III).
SQ SEQUENCE 1468 AA; 163809 MW; EF2FBIE9CA9DAF43 CRC64;

Query Match 4.0%; Score 110; DB 1; Length 1468;
Best Local Similarity 17.1%; Pred. No. 11;
Matches 98; Conservative 73; Mismatches 185; Indels 216; Gaps 21;
QY 9 KRWLPLEVPTEPTLGHRLSLRLCTWGYSSNTRFTITLNYKDPLTGDEPTL-----63
Db 517 KRLLSPVSP-LSPRCPSPQHQHRIILL-----PLPSEGEVWFNEYP 557
QY 64 -----ASYGIVSGDLICILLODDIPAPNIPST-----DSEHSSIQNNEQPSSLATSSN 111
Db 558 SRKNDVSSGFPSSADTL-----EPSSTTKVTETKGASPTSLRASQTMLVSEAS 605
QY 112 QTSMDQEQ-----PSDSFOGAAO-----SGVNDSDSLGCPSONFE 147
Db 606 EKLGPKEITAPQHQHLPPIGASEGFCDFKTEATKDLNPKDGVWVPGYRAGPPCPFL 665
QY 148 AESTQDNAHMAEGTFYPSEPMLCSEVGEQVPHSLTLYQSADCSANDALIVLIHLLM 207
Db 666 LHEEKETSRSE--LYLDLNPQDSPTEDDRTPCRLQAVWPPPKTKDTEEK---VGLKY 719
QY 208 LESGYIQGTAKALSMPEKWK-----LSGVYKLOYHPLCEGSSATLTCVPLGNLIVN 262
Db 720 TEAEY-----QAATLHLKREHKEIETLQAOFLKTFH--IRGEHALVTA-----762
QY 263 ATLKINNEIRSVKRLQLLPESFICKELGENVANIYDLOKLRLFKDQLVYPLLAFTQQ 322
Db 763 -----RLEAENLKQLEKRRREGCEM-----785
QY 323 ALNLPDVGVLVPLLEKLKLRIFRLDVRSLSLSAVCARDLFTASND---PLLWRFYLRD 379
Db 786 -----RDVCISDTDDCSPKAFRNVCIQT 808
QY 380 FRDNTVRVQDQDWKELYRKRHIQRK-----ESPQGRF-----VML 414
Db 809 DRETFLKPCAESKATRSSQIVPKKLTISLTQSPSKDSKDIHAPFQTRGTSSSQOKI 868
QY 415 LPSSTHTIYPNPLHPRPPSSRLPGIIGCEYDQRTLPYVGDPISSLIPGGEFSPQ 474
Db 869 SPAPAPPPLPPLPILPPPPP--LPEGL--GPLPPAPPPIP-----PVCVSPSPPPPPPP 919
QY 475 FPLLRPRFDPVGPLPQGNPI-----LFCRGGP 501
Db 920 PTPVPPSDGPPPPPPPPPLPNVLPNSGGP 951

```

RESULT 14
ID NFI_HUMAN STANDARD; PRT; 772 AA.
AC Q1494; Q12877;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear factor erythroid 2 related factor 1 (NF-E2 related factor 1)
DE (NFE2-related factor 1) (Nuclear factor, erythroid derived 2, like 1)
DE (Transcription factor 11) (Transcription factor HBZ17) (Transcription
DE factor LCR-Fl) (Locus control region-factor 1).
GN NFE2L1 OR NFE1 OR TCF11 OR HBZ17.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95095252; PubMed=8001966;
RA Luna L., Johnson O., Skartlien A.H., Pedetour F., Turc-Carel C.,
RA Prydz H., Koistoe A.-B.;
RT "Molecular cloning of a putative novel human bZIP transcription
RT factor on chromosome 17q22.";
RT Genomics 22:553-562(1994).
RN [2]
RP SEQUENCE OF 326-772 FROM N.A.
RX MEDLINE=94310089; PubMed=8036168;
RA Caterina J.J., Donze D., Sun C.W., Clavatta D.J., Townes T.M.;
RT "Cloning and functional characterization of LCR-Fl: a bZIP
RT transcription factor that activates erythroid-specific, human globin
RT gene expression.";
RL Nucleic Acids Res. 22:2383-2391(1994).
CC -1- FUNCTION: ACTIVATES ERYTHROID-SPECIFIC, GLOBIN GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. CNC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X77366; CAA54555.1; -
DR EMBL; U08853; AAA20466.1; -
DR HSP; P34707; ISKN.
DR MIM; 163260; -
DR InterPro; IPR001871; bZIP.
DR Pfam; PF001170; bZIP; 1.
DR SMART; SM00338; BRZ; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT DOMAIN 125 288 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 496 517 POLY-SER.
FT DOMAIN 659 674 BASIC MOTIF.
FT DOMAIN 682 704 LEUCINE-ZIPPER.
SQ SEQUENCE 772 AA; 84703 MW; C868807C6045BEF5 CRC64;
Query Match 4.0%; Score 109.5; DB 1; Length 772;
Best Local Similarity 20.8%; Pred. No. 5;
Matches 80; Conservative 56; Mismatches 139; Indels 109; Gaps 17;
Qy 17 PETE-----PTLGLRSHRLRLSLCTGYSSNTR--FTITLN-----YKDPIT 57
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 PEVSLPVASSTLLPLAPSNSTSLNFTSGTNTLGLFFPPQLNGTANDTAGPELPDPLG 439
Yy 58 G-----DETLASGYGVSGDLICLILODDIPAPNPSSDTSEHS-SLQNEOP-SLAT 108
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 GLLDEAMIDETSLMDLATEEG-----FNPQVASQLEEFDSGSLDSHSPSLSS 492
Yy 109 SSNOTSMQDEQPSDFQCAQAQSGVWDDSMGLGSPNQFAESIODNHAHEG-TGFYFSE 167

```

```
DR MGD; MGI:99421; Nfe211.
DR InterPro: IPR001871; bZIP.
DR Pfam: PF00170; bZIP; 1.
DR SMART: SM00338; BRLZ; 1.
DR DR PROSITE: PS00036; bZIP_BASIC; 1.
KW DNA-binding; Nuclear protein; Phosphorylation; Alternative splicing.
FT DOMAIN 125 259
FT DOMAIN 414 447
FT DOMAIN 467 486
FT DOMAIN 628 643
FT DOMAIN 651 673
FT DOMAIN 730 737
FT VARSPIC 1 291
FT VARSPIC 447 583
FT CONFLICT 318 318.
FT CONFLICT 387 387
SQ SEQUENCE 741 AA; 81545 MW; C01B89DD26E7CDEF CRC64;

Query Match      . 4.0%; Score 109; DB 1; Length 741;
Best Local Similarity 20.5%; Pred. No. 5.1;
Matches 67; Conservative 51; Mismatches 119; Indels 90; Gaps 14;

QY 54 DPLTG-----DEETLASYGVSGDLICLILQDDIPAPNIPSTDSHS-SLQNEQPS 105
DB 407 DPLGGLDEAMLDEISLMDLAIEG-----FNPVQASQLEEFDSGLSDSSHSPS 459
QY 106 LATSSNOTSQDEQSDSFQQAQSGVWDDSMGLPSONFEAESIQDNAMHAEQ-TGFY 164
DB 460 SLSSSEGSSSSSSSSSSSASSS-FSEEGAVGYSDSETLLEE---AEGAVGYQ 514
QY 165 PSEPMLCSEVGVPHSLTEL-----YQADCSNDALIVLHLMLESGYI 213
DB 515 PEYSKFCRMSY--QDPSQLSCLPYLEHVGHNTYMAPSA-----LDSADL 558
QY 214 P-----QGTEAKALSMPEKWKLSGVYKLOYMHPLCEGSSATLTCVPLGNLIVVNATLKI 267
DB 559 PPPSTLKGSKKQADFLDKQMSRDEHRAAMK-----IPFTNDKIINLPVEE 606
QY 268 NNEIRSVKRLQLLPESFI-----CKEKLGENVANIYKDLQKLSRLFKDQL 312
DB 607 FNELLSKVQLSEAGLSLIRDIRRGKNKMAAQNCRRKKLDTILNLERDVEDLQR-DKARL 665
QY 313 VYPLLATFTRQALNL-----PDVFG 331
DB 666 LREKVEFLRSLRQMKQKVQSLYQEVFG 692
```

Search completed: June 6, 2002, 11:34:05
Job time: 219 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2002, 11:29:56 ; Search time 30.83 Seconds
(without alignments)
2929.074 Million cell updates/sec

Title: US-09-927-458-2
Perfect score: 2754
Sequence: 1 MRLRVLLKRTWPLEVPETE.....DRFPFRSGRPTDGLRSLFM 522

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	2744	99.6	522	Q96HM6	Q96hm6 homo sapien
2	386	14.0	75	Q9UH23	Q9uh23 homo sapien
3	240.5	8.7	475	Q9ZUB8	Q9zub8 arabidopsis
4	187.5	6.8	350	Q9ZUB9	Q9zub9 arabidopsis
5	145.5	5.3	434	Q9Y593	Q9y593 homo sapien
6	145.5	5.3	547	Q9NT57	Q9nt57 homo sapien
7	144	5.2	387	Q96CU6	Q96cu6 homo sapien
8	128.5	4.7	850	Q17055	Q17055 caenorhabdi
9	126	4.6	355	Q96P95	Q96p95 homo sapien
10	120.5	4.4	2158	Q9WU13	Q9wu13 rattus norv
11	120	4.4	1255	Q9R2J6	Q9r2j6 shigella so
12	119	4.3	287	Q95UG5	Q95ug5 babesia bov
13	119	4.3	844	Q9R2J5	Q9r2j5 shigella so
14	118	4.3	442	Q9VG61	Q9vg61 drosophila
15	118	4.3	584	O53539	O53539 mycobacteri
16	117.5	4.3	1179	Q9NQW1	Q9nqw1 homo sapien

17	117	4.2	2087	11	Q9WUE8	Q9wue8 rattus norv
18	117	4.2	2167	11	Q9WV48	Q9wv48 rattus norv
19	116.5	4.2	915	4	Q9Y4O8	Q9y4o8 homo sapien
20	116	4.2	4957	4	O14687	O14687 homo sapien
21	116	4.2	5262	4	O14886	O14886 homo sapien
22	115.5	4.2	355	11	Q9CPU7	Q9cpu7 mus musculu
23	115.5	4.2	2161	4	Q9Y566	Q9y566 homo sapien
24	114.5	4.2	2360	3	O94188	O94188 cryptococcu
25	114	4.1	245	4	Q96BG5	Q96bg5 homo sapien
26	113	4.1	326	2	Q9RFW8	Q9rfw8 mycoplasma
27	113	4.1	1337	4	Q96RR2	Q96rr2 homo sapien
28	113	4.1	1338	4	Q9BZG0	Q9bzg0 homo sapien
29	113	4.1	1338	4	Q96RG4	Q96rg4 homo sapien
30	113	4.1	1339	4	Q96RG5	Q96rg5 homo sapien
31	112.5	4.1	350	11	Q91Z62	Q91z62 rattus norv
32	112	4.1	1006	10	Q9LWQ1	Q9lqw1 arabidopsis
33	111.5	4.0	655	4	Q96Q01	Q96q01 homo sapien
34	111.5	4.0	940	4	O00405	O00405 homo sapien
35	111.5	4.0	940	4	Q96QC0	Q96qc0 homo sapien
36	111	4.0	463	4	Q9BRQ0	Q9brq0 homo sapien
37	111	4.0	1097	10	Q9LY69	Q9ly69 arabidopsis
38	111	4.0	1427	13	Q9DG29	Q9dg29 xenopus lae
39	110.5	4.0	551	4	Q16630	Q16630 homo sapien
40	110.5	4.0	602	2	Q9AKP3	Q9akp3 rickettsia
41	110.5	4.0	804	10	Q94HX0	Q94hx0 oryza sativ
42	110.5	4.0	1315	10	Q9SPM0	Q9spm0 zea mays (m
43	110	4.0	288	10	Q9AVH4	Q9avh4 pisum sativ
44	109.5	4.0	389	11	O88539	O88539 mus musculu
45	109.5	4.0	453	2	Q9ADR3	Q9adr3 staphylococ

ALIGNMENTS

RESULT 1

Q96HM6 ID Q96HM6 PRELIMINARY; PRT; 522 AA.
AC Q96HM6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F-BOX ONLY PROTEIN 7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREATIC ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008361; AAH08361.1;
SQ SEQUENCE 522 AA; 58484 MW; B3702EE5C2149504 CRC64;

Query Match 99.6%; Score 2744; DB 4; Length 522;
Best Local Similarity 99.6%; Pred. No. 2.7e-230;
Matches 520; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MRLRVLLKRTWPLEVPETEFTLGHLSHLRLSLCTWGYSSNTRFTTLNKKDPLTGDE	60
Db	1	MRLRVLLKRTWPLEVPETEFTLGHLSHLRLSLCTWGYSSNTRFTTLNKKDPLTGDE	60
Qy	61	ETLASYGVISGDLICLLIQDDIPAPNIPSSDSEHSSLQNNQPSLATSSNOTSMODROP	120
Db	61	ETLASYGVISGDLICLLIQDDIPAPNIPSSDSEHSSLQNNQPSLATSSNOTSIQDSQP	120
Qy	121	SDSFQGAQAQSGVNDSDMLGFSQNFQAEQSDQNAHMAEGTGFYFSEPMLCSEVGEQVP	180
Db	121	SDSFQGAQAQSGVNDSDMLGFSQNFQAEQSDQNAHMAEGTGFYFSEPMLCSEVGEQVP	180
Qy	181	HSLETLYOSADCSANDALIVLIHLLMLESYIPQGTAKALSMPEKWKLSGVYKQYMH	240

Db	181	HSLETYSADCSANDALIVLIHLLMESGYIPQTEAKALSMPEKWKLSGVYKLOYMH	240
Qy	241	PLCESSATLTCVPLGNLIVVNATLKINNEIRSVKRLQLLPESFICKKLGENVANIYKD	300
Db	241	PLCESSATLTCVPLGNLIVVNATLKINNEIRSVKRLQLLPESFICKKLGENVANIYKD	300
Qy	301	LQKLSRLFKDQVYVLLAFTQAALNLPDVGVLVLPLEKLRIFRLDVRVLSLAVCR	360
Db	301	LQKLSRLFKDQVYVLLAFTQAALNLPDVGVLVLPLEKLRIFRLDVRVLSLAVCR	360
Qy	361	DLFTASNDPLLRFLYLRDRDNTVRVODTWKELYRKRHRHQRKESPKGRVMLLPSTH	420
Db	361	DLFTASNDPLLRFLYLRDRDNTVRVODTWKELYRKRHRHQRKESPKGRVMLLPSTH	420
Qy	421	TIPFPYPLHPRFPSSRLPPGIIGGEYDQRPPTLPYVGDPTSSLLPGPGETPSQFPPLRP	480
Db	421	TIPFPYPLHPRFPSSRLPPGIIGGEYDQRPPTLPYVGDPTSSLLPGPGETPSQFPPLRP	480
Qy	481	REDVPGPLGPNLPILPGRGPNDRFPFRSRRPTDGRLESEM	522
Db	481	REDVPGPLGPNLPILPGRGPNDRFPFRSRRPTDGRLESEM	522
RESULT	2		
Q9UH23		PRELIMINARY;	PRT; 75 AA.
AC	Q9UH23;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	DJ116G19.1 (NOVEL PROTEIN) (FRAGMENT).		
GN	CN28H9.1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Clark G.;		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL035068; CAB41542.1;		
FT	NON_TER 1		
FT	NON_TER 75		
FT	NON_TER 75		
SEQ	SEQUENCE 75 AA; 8346 MW; 73224AEF1BB1FC4A CRC64;		
Query Match	14.0%;	Score 386;	DB 4; Length 75;
Best Local Similarity	100.0%;	Pred. No. 3e-26;	Indels 0; Gaps 0;
Matches	75; Conservative	0; Mismatches	0;
Qy	216	GTEAKALSMPEKWKLSGVYKLOYMHPLCEGSSATLTCVPLGNLIVVNATLKINNEIRSVK	275
Db	1	GTEAKALSMPEKWKLSGVYKLOYMHPLCEGSSATLTCVPLGNLIVVNATLKINNEIRSVK	60
Qy	276	RLQLLPESFICKEL 290	
Db	61	RLQLLPESFICKEL 75	
RESULT	3		
Q9ZUB8		PRELIMINARY;	PRT; 475 AA.
AC	Q9ZUB8;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	F508.33 PROTEIN (AF1G23780/F508_31).		
GN	F508.33.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		

RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=CV. COLUMBIA;		
RC	Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Liu S., Lenz C.,		
RA	Li J., Kremenetskaia I., Luros J., Altafi H., Gonzalez A., Araujo R.,		
RA	Buehler E., Conn L., Conway A.B., Dunn P., Hansen N., Huizar L.,		
RA	Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,		
RA	Ecker J.R., Federpiel N.A., Theologis A.;		
RT	"Arabidopsis thaliana chromosome 1 BAC F508 sequence.";		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	STRAIN=CV. COLUMBIA;		
RC	Theologis A.;		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,		
RA	Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,		
RA	Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,		
RA	Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,		
RA	Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,		
RA	Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,		
RA	Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,		
RA	Ecker J.R.;		
RT	"Arabidopsis cDNA clones.";		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AC005990; AAC98033.1;		
DR	EMBL; AY049291; AAK83633.1;		
DR	InterPro; IPR001810; F-box.		
DR	InterPro; IPR002052; N6_Mtase.		
DR	Pfam; PF00646; F-box; 1.		
DR	SMART; SM00256; FBOX; 1.		
DR	PROSITE; PS50181; FBOX; 1.		
DR	PROSITE; PS00092; N6_MTASE; UNKNOWN_1.		
SEQ	SEQUENCE 475 AA; 52947 MW; D068CCF15448EFC3 CRC64;		
Query Match	8.7%;	Score 240.5;	DB 10; Length 475;
Best Local Similarity	24.5%;	Pred. No. 2.1e-12;	
Matches	116; Conservative	89; Mismatches	177; Indels 91; Gaps 20;
Qy	1	MRLRVLLKRTWPLEVTEPTLHGLRSLRLSLCTWGYSSNTRFTTLNKKDPL--TG 58	
Db	3	LRLRHETRETLLKLELADAD-TLHLRRRINPTV-----PSSVHLNARKDELITPS 53	
Qy	59	DEETLASYGVSGDLICLILQDDIPAPNIPSTDSEHSLQNNQPSLATSSNOTSMODE 118	
Db	54	PEDTLRLSLGLISGLIYFSLE-----AGESNNKLRDSETVASQSSNQTSVHD- 102	
Qy	119	QPSDSFGQAAGSGVYNDSDMLGSPQNFEAESIQDNAHMAEG--TGFPYSEPM----- 169	
Db	103	--SIGF-----AEVDVVPQAKSNPN-----TSVED-----PEGDISGMEGPMDVEQLDM 147	
Qy	170	--LCSESVEGQVPHSLET--LYQADCDANDALIVLIHLLMESGYI--PGTE----- 218	
Db	148	ELAAAGSKRLSEPFPLKNILLEKSGDTSBELT-TLALSHAVMLSEGFVLLNHGSKFNFS 206	
Qy	219	AKALSMPEKWKLSGVYKLOYMHPLCEGSSATLTCVPLGNLIVVNATL-----KINNEI-- 271	
Db	207	KELLTVSLRYTLPelikSKDTNTI---ESVSVKFNQLGPPVVVYGVTVGSSGRVHNLDK 263	
Qy	272	-RSVKRLQLLPESFICKKLGENVANIYKDLQKLSRLFKDQVYVLLAFTQAALNLPDVF 330	
Db	264	RRVPVVDLVMDDTSTDSDE--EGSSSIYREVFMFWMVKDRLVPIILLICIDKAGLEPPP 320	
Qy	331	GLVVLPLEKLRIFRLDVRVLSLAVCRDLFTASNDPLLRFLYLRDRFRONTVRV--Q 388	
Db	321	CLMRPLELKLKLELLPGVSGNMACVCTEMRYLSTANDLNRKQKCLEEVNFVTEAGD 380	
Qy	389	DTQWK-----ELYKRKHRTQKESPKGRVMLLPSTHTIPF 424	
Db	381	SVNWKARFATFWKQKQLAAASDTFWQNLGRNISTGRSGIRFPRIIGDPPF 433	


```
RESULT 4
Q9ZUB9 PRELIMINARY; PRT; 350 AA.
AC Q9ZUB9
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F508.32 PROTEIN.
GN F508.32
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Liu S., Lenz C.,
RA Li J., Kremetska I., Luros J., Altafi H., Gonzalez A., Araujo R.,
RA Buehler E., Conn L., Conway A.B., Dunn P., Hansen N., Huizar L.,
RA Kim C., Palm C.J., Rowley D., Shinn P., Walker M., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RA "Arabidopsis thaliana chromosome 1 BAC F508 sequence.";
RT Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005990; AAC98032.1;
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PS0181; FBOX; 1.
SQ SEQUENCE 350 AA; 39953 MW; CD0BD4C6791674D6 CRC64;

Query Match 6.8%; Score 187.5; DB 10; Length 350;
Best Local Similarity 24.4%; Pred. No. 5.5e-08;
Matches 90; Conservative 58; Mismatches 118; Indels 103; Gaps 17;

Qy 172 SESVEGVPHSL-----TLQSADCSANDALIVLHLLMLESYI 213
Db 22 NSGIEGVPMDVELAAAKSKRLSEPFPLKNVLEKSGDTSALT-ALASVHVMLESGFV 80
Qy 214 --PGTE-----AKALSMPEKWLKSGYKQYMHPLCEGSSATLTCVPLGNLVNATL- 265
Db 81 LLDHGSDFKFSKLLSVLRITLPELITRKDTNTV---ESVTVRFONIGRLVYVYGLT 137
Qy 266 ----KINNEIRSVKRLQLLPESFICKEL---GENVANIYKDLKSLRFLKDLQVYPLLA 318
Db 138 GSKCRVH--WTSLDKSRFLPVIDLVDTLKPEKQSSSYREVFELMRVMDLVIPLL- 194
Qy 319 FTRQALNLPDVF-----LVVLEPLKLRIFRLDVRVLSLSAVCRDLFTASNDPLW 372
Db 195 -----IGLCKAGLESPPCLMLLTTELKLTLELPGVIGYMACVCTEMRYLASNDLW 249
Qy 373 RFLYLRDFRONTVRVQ--DIDWKELY-----RKR--HQKESPKGFVMLLPSTHTIP 423
Db 250 EHKCLEGKGLWLYTGVDWKRKFKASFWRKRKLDLLIARNPPITK----- 296
Qy 424 FYPNLPFRPPSSRLPPGIIGGYDQRTPLPYVGDPISLIPGPGTSPFPPL----- 478
Db 297 --SNPFRFTLPDR-----DRREPDRFG-----PSDFYRFLRDP 331
Qy 479 RPRFDPVGP 487
Db 332 RDRFGPRDP 340

RESULT 5
Q9NT57 PRELIMINARY; PRT; 547 AA.
AC Q9NT57
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

Q9Y593 PRELIMINARY; PRT; 434 AA.
AC Q9Y593
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NY-REN-57 ANTIGEN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99438124; PubMed=10508479;
RA Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA Jongeneel V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA Old L.J.;
RT "Antigens recognized by autologous antibody in patients with renal-
RT cell carcinoma.";
RL Int. J. Cancer 83:456-464(1999).
DR EMBL; AF155114; AAD42880.1;
DR InterPro; IPR001810; F-box.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00646; F-box; 1.
DR Pfam; PF00515; TPR; 1.
DR PROSITE; PS0181; FBOX; 1.
FT NON_TER
SQ SEQUENCE 434 AA; 48863 MW; CF2A074E8371120D CRC64;

Query Match 5.3%; Score 145.5; DB 4; Length 434;
Best Local Similarity 21.7%; Pred. No. 0.00034;
Matches 98; Conservative 61; Mismatches 123; Indels 169; Gaps 24;

Qy 105 SLATSSNOTSMQDEQPSDFOGAAQSCVWDDSMLG-PSONFAESIQDQNAHAEGTGF 163
Db 75 SSAETGCRSRHPDTHPSS--GGRCRGGTSPSSAAGRPASMAEAE-----EDCH----- 122
Qy 164 YPSEPMLCSSEVGQVPHSLTLYQSADCSANDALIVLHLLMLESYI 222
Db 123 -----SDIVR-----ADDEENES-----PAETDLQAL 146
Qy 223 SMPEKWLKSGYKQYMHPLCEG--SSATLTCVPL-----GNLIVNATLKINNE----- 270
Db 147 QM-----FRAQWMPFELAPGVSSSNLENRPCRAARGSLQKTSADTKGQEQAKEKA 197
Qy 271 ----IRSV-----KRLQLPE-----SPICEKLGENVANIY---KDQLQ 302
Db 198 RELFLKAVEEQNGALYEAIFKRYRAMOLVPDIEFKITYTRSPDGDGVGNSYIEDNDDDS 257
Qy 303 KLSRL---FKDQLVYPLLAFTROALNL--PDV---FGLVVLPLELKLIRFL-----LD 348
Db 258 KNADLLSYFQQQ-----LTFQESVLKLCQPELESSQIHISVLPMEVLMIFRWVSSDL 312
Qy 349 VRSVLSLSAVCRDLFTASNDPLLRFLYLRDFRONTVR-VQDTWKELYRKRHHQKRESP 407
Db 313 LRSLEQLSLVGRGYICARDPEIWLRLACKLVGRSKICLVPTYSWREMFLEP-----P 365
Qy 408 KGRFVMLLPST-----HTIPEYPNLHPRPPSPSSRLPPGIIGGYDQ 450
Db 366 RVREDGVYISKTTIROGEQLDGFYRAWHQVEY---RYIRFFPDGHV----- 411
Qy 451 RPTLPYVGDPISLIPGPGTSPFPPLRPR 481
Db 412 -----MMLTTPPEQSIPLRLTR 430

RESULT 6
Q9NT57 PRELIMINARY; PRT; 547 AA.
AC Q9NT57
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
```

DT	01-OCT-2001 (TReMBLrel. 18, Last annotation update)	
DE	HYPOTHETICAL 62.2 KDA PROTEIN (FRAGMENT).	
GN	DKEZP434C0118.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=TESTIS;	
RA	Ottensmeyer B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;	
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AL137520; CAB70786.1;	
DR	InterPro; IPR001810; F-box.	
DR	InterPro; IPR001440; TPR.	
DR	pfam; PF00646; F-box; 1.	
DR	PROSITE; PS0181; FBOX; 1.	
KW	Hypothetical protein.	
FT	NON_TER	
SQ	SEQUENCE 547 AA; 62217 MW; A726E815A82DF31 CRC64;	
Query Match 5.3%; Score 145.5; DB 4; Length 547;		
Best local similarity 21.7%; Pred. No. 0.00048;		
Matches 98; Conservative 61; Mismatches 123; Indels 169; Gaps 24;		
QY	105 SLATSNQTSMDQEQSDSFQGAAGGVNDSDMLG-PSQNFEEESIODNAHMAEGTGF 163	
DB	72 SAAETGRSHPTQHPSS--GGRCGGTSPSSAAGRFAAEEA--EDCH----- 119	
QY	164 YPEPMLCSSEVGGVPHSLQSDGSDANDALIVLHLMLESGYIPQGTAKA-L 222	
DB	120 -----SDTVR-----ADDEENES-----PAETDLQAOL 143	
QY	223 SMPEKWLGVYKLYMHPICEG-SSATLTCVPL-----GNLIVVNATLKINNE----- 270	
DB	144 QM-----FRAQWFMELAPGVSSNLENPCRAARGLQTSADTKGQEQAKEKA 194	
QY	271 -----IRSV-----KRLQLPE-----SFICKEKLGENVANIY-----KDLQ 302	
DB	195 RELFKAVEEENGALYEAIKFYRRAMQLVPDIEFKITYTRSPDGVGNSYIEDNDDSD 254	
QY	303 KLSRL-----FKDQLVPLLATRAALNL--PDV-----FGLVVLPLELKLRIPL 348	
DB	255 KMDLLSVFQQQ-----LTFQESVLKLCQPELESSQIHISVLPMEVLMYIFRWVSSDLD 309	
QY	349 VRSVLSLAVCRDLFTASNDPLLRFLYLRDRFNTVR-VQDQWKELYKRKHQKES 407	
DB	310 LRSLEQLSVCRGYICARDPEIWRACLKVGGRSCIKLVPYTSWREMFLEP-----P 362	
QY	408 KGRFVMLLPSS-----HTIPFPNPLHPRFPFSSRLPPGIIGGYDQ 450	
DB	363 RVREDGVYISKTVIRQEQSLDGFYRAWHQVEYI---RYIRFPDGHV----- 408	
QY	451 RPTLPYVGDPSLIPGGETSPQFPPLRPR 481	
DB	409 -----MMLTTPPEQSIQVPLRTR 427	
RESULT 7		
Q96CU6	PRELIMINARY; PRT; 387 AA.	
AC	Q96CU6;	
DT	01-DEC-2001 (TReMBLrel. 19, Created)	
DT	01-DEC-2001 (TReMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	
DE	HYPOTHETICAL 45.4 KDA PROTEIN.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=TESTIS;	
RA	Ottensmeyer B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;	
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AL137520; CAB70786.1;	
DR	InterPro; IPR001810; F-box.	
DR	InterPro; IPR001440; TPR.	
DR	pfam; PF00646; F-box; 1.	
DR	PROSITE; PS0181; FBOX; 1.	
KW	Hypothetical protein.	
FT	NON_TER	
SQ	SEQUENCE 547 AA; 62217 MW; A726E815A82DF31 CRC64;	
Query Match 5.2%; Score 144; DB 4; Length 387;		
Best local similarity 23.1%; Pred. No. 0.00039;		
Matches 74; Conservative 47; Mismatches 90; Indels 110; Gaps 16;		
QY	233 VYKLYMHPICEG-SSATLTCVPL-----GNLIVVNATLKINNE-----IRSV-- 274	
DB	35 MFRQWFMELAPGVSSNLENPCRAARGLQTSADTKGQEQAKEKARELFLKAVEE 94	
QY	275 -----KRLQLPE-----SFICKEKLGENVANIY-----KDLQKLSRL---FK 309	
DB	95 EONGALYEAIKFYRRAMQLVPDIEFKITYTRSPDGVGNSYIEDNDDSKMDLLSVFQ 154	
QY	310 DQLVPLLATRAALNL--PDV-----FGLVVLPLELKLRIPL-----LQVRSVLSLV 358	
DB	155 QQ-----LTFQESVLKLCQPELESSQIHISVLPMEVLMYIFRWVSSDLDLRSLEQLSLV 209	
QY	359 CRDLFTASNDPLLRFLYLRDRFNTVR-VQDQWKELYKRKHQKESPKGRFVMLLP 417	
DB	210 CRGFYICARDPEIWRACLKVGGRSCIKLVPYTSWREMFLEP-----PVRFDGVVIS 262	
QY	418 ST-----HTIPFPNPLHPRFPFSSRLPPGIIGGYDQRTLPYVGD 460	
DB	263 KTYIRQEQSLDGFYRAWHQVEYI---RYIRFPDGHV----- 298	
QY	461 ISSLIPGGETSPQFPPLRPR 481	
DB	299 -----MMLTTPPEQSIQVPLRTR 317	
RESULT 8		
O17055	PRELIMINARY; PRT; 850 AA.	
AC	O17055;	
DT	01-JAN-1998 (TReMBLrel. 05, Created)	
DT	01-OCT-2001 (TReMBLrel. 18, Last sequence update)	
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	
DE	HYPOTHETICAL 93.2 KDA PROTEIN.	
GN	C24A1.3.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BRISTOL N2;	
RX	MEDLINE=99069613; PubMed=9851916;	
RA	None;	
RT	"Genome sequence of the nematode C. elegans: a platform for	
RT	investigating biology. The C. elegans Sequencing Consortium."	
RL	Science 282:2012-2018(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BRISTOL N2;	
RA	Connell M.;	
RT	"The sequence of C. elegans cosmid C24A1.3;	
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BRISTOL N2;	
RA	Waterston R.;	
RT	"Direct Submission."	
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF024491; AAB70312.2;	
DR	HSSP; P42773; 1IIB	
DR	InterPro; IPR002110; ANK.	

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lakso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy L., Murphy B., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AF003696; RAF54827.1;
DR FlyBase; FBgn0038056; CG5961.
DR InterPro; IPR001810; F-box.
DR DR InterPro; IPR001440; TPR.
DR DR Pfam; PF00646; F-box; 1.
DR DR Pfam; PF00515; TPR; 1.
DR DR PROSITE; PS0181; FBOX; 1.
SQ SEQUENCE 442 AA; 50963 MW; FE6CE011AD5E09C5 CRC64;

Query Match 4.3%; Score 118; DB 5; Length 442;

Best Local Similarity 27.3%; Pred. No. 0.087;

Matches 42; Conservative 32; Mismatches 56; Indels 24; Gaps 8;

QY 266 KINNEIRSVKRLQ---LPESFCKEKLGENVA-NIYKDLQLSLFKDQLYV--PLLA 319
Db 111 KLSNDV-SKKYLNLANLAKQLDGSODGEEVVNDLY---EKQHDLRKNNKMIAS 166
QY 320 TROALNLPDVGIVLPLELKLRIFR-----LLDVRSVLSAVCRDLFTASNDPLLWRF 374
Db 167 SRDANVLTGLHPADLPPEIVMRILRWVSAQLDMRSLEQCAVCKGFVYARDELWRL 226
QY 375 LYLRDRFDN--TVRQDFTD-----WKELYRKR 399
Db 227 ACVKVGHNVHTLEAQSDSVSNVHFHSWRDMFIRR 260

RESULT 15

OS3539

ID 053539 PRELIMINARY; PRT; 564 AA.

AC 053539;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 59.6 KDA PROTEIN.

GN RV3494C OR MT023.01C.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: AL022022; CAAL7731.1;
DR TubercuList; RV3494c;
DR InterPro; IPR003399; Mce.
DR Pfam; PF02470; mce; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 564 AA; 59636 MW; 37752501FE15EC6F CRC64;

Query Match 4.3%; Score 118; DB 16; Length 564;

Best Local Similarity 25.5%; Pred. No. 0.12;

Matches 51; Conservative 18; Mismatches 51; Indels 80; Gaps 11;

QY 354 SLSAVCRDLF--TASNDPLLWRLFLYLRDRONTVR-VQDTDWKELYRKR--HIQRKESPK 408
Db 343 SVREIPRDMYCKTAQNDP-----STVRGARNYPCQEPFGKRAPTVQLCRDPR 389
QY 409 GRFVMLPSSTH-----TIPFYP-----NPLHPRPPFPSSRLPGIIGGEYDQRFTLPYV 457
Db 390 G-----YVPVGTNPWRGPDIPYGTETDGRNLLPNNKFP--YIPPGA-----DPDGPVPIV 438
QY 458 GDPTSSLLPFGGETPSQ-----FPLLRPKF----- 482
Db 439 GPPPPGVAGGPAHPAQAPAPPNDNGPPPPFTSWMPGYPPEPPQVYPYPIATIPPPPP 498
QY 483 -----DPVGPLPGPNPILPG 497
Db 499 PEGTGPFGPAPGPOPOQASG 518

Search completed: June 6, 2002, 11:33:45

Job time: 229 sec